



# **STIC Search Report**

## **Biotech-Chem Library**

STIC Database Tracking Number: 116176

TO: Phillip Gambel  
Location: 3e81/3c70  
Tuesday, March 09, 2004  
Art Unit: 1644  
Phone: 272-0844  
Serial Number: 10 / 042421

From: Jan Delaval  
Location: Biotech-Chem Library  
Rem 1A51  
Phone: 272-2504

[jan.delaval@uspto.gov](mailto:jan.delaval@uspto.gov)

### Search Notes

10/042421

## SEARCH REQUEST FORM

116176

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_

Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

### Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

### STAFF USE ONLY

Date completed: 3/9/04

Searcher: Jan

Terminal time: \_\_\_\_\_

Elapsed time: 12 + 10

CPU time: \_\_\_\_\_

Total time: \_\_\_\_\_

Number of Searches: \_\_\_\_\_

Number of Databases: \_\_\_\_\_

#### Search Site

☒ STIC

☐ CM-1

☐ Pre-S

#### Type of Search

☐ N.A. Sequence

☒ A.A. Sequence

☐ Structure

☐ Bibliographic

#### Vendors

☐ IG

☐ STN

☐ Dialog

☐ APS

☐ Geninfo

☐ SDC

☒ DARC/Questel

☐ Other

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

## OM protein - protein search, using sw model

Run on: March 8, 2004, 06:08:16 ; Search time 21 Seconds

(without alignment)  
2258.211 Million cell updates/sec

Title: PCT-US01-51014-1

Sequence: 1 MDKFWHAAWGLCLVPLSLA.....QFMADETRNLCNVDMKIGV 493

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78: \*  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2590	99.3	493	2	CD44E protein, epi
2	2490	95.5	699	2	epican - human
3	2468.5	94.7	742	2	lymphocyte homing
4	2261	86.7	426	2	lymphocyte homing
5	2036	78.1	395	2	CD44R5 - human
6	1835	70.4	351	2	cell adhesion mole
7	1698.5	65.1	362	2	lymphocyte adhesio
8	1547	59.3	359	2	lymphocyte surface
9	1470	56.4	351	2	CD44 antigen precu
10	1469.5	56.3	366	2	cell-surface glyco
11	1427.5	54.7	362	2	T-cell surface gly
12	1425	54.6	503	2	cell adhesion mole
13	1368	52.5	363	2	CD44 homolog membr
14	1361	52.2	365	2	CD44 membrane gly
15	699	26.8	337	2	CD44 protein - hum
16	541.5	20.8	146	2	cell surface glyco
17	525	20.1	148	2	cell surface glyco
18	394	15.1	301	2	CD44 glycoprotein
19	382.5	14.7	142	2	CD44 glycoprotein
20	380	14.6	414	2	CD44 protein - mou
21	344	13.2	70	2	cell surface glyco
22	233.5	9.0	107	2	CD44 glycoprotein
23	183	7.0	2232	2	hypothetical prote
24	178	6.8	833	2	lipoprotein vac
25	171	6.6	276	2	TSG-6 homolog PS4
26	168	6.4	2124	2	proteoglycan core
27	165	6.3	786	2	hypothetical prote
28	163	6.2	275	2	tumor necrosis fac
29	162.5	6.2	277	2	hyaluronate-bindin

30	161.5	6.2	631	2	epistatin - mouse
31	161	6.2	1275	2	hypothetical prote
32	158	6.1	630	2	tumor-associated m
33	156.5	6.0	1459	2	hypothetical prote
34	156	6.0	1032	2	hypothetical prote
35	155.5	6.0	2132	2	aggreacan precursor
36	152.5	5.8	1643	2	aggreacan precursor
37	152.5	5.8	3381	2	versican precursor
38	152.5	5.8	2109	1	aggreacan precursor
39	150.5	5.8	636	2	probable membrane
40	150	5.8	862	2	versican - pig-tai
41	150	5.8	888	2	secreted acid phos
42	149	5.7	1217	2	sericidinb - silkw
43	148.5	5.7	2271	2	hypothetical prote
44	148.5	5.7	2397	1	versican precursor
45	148	5.7	1367	1	glucan 1,4-alpha-g

## ALIGNMENTS

RESULT 1									
S13530									
CD44E protein, epithelial - human									
C/Species: Homo sapiens (man)									
C/Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000									
C/Accession: S13530									
R/Stamenkovic, I.; Aruffo, A.; Amiot, M.; Seed, B.									
EMBO J. 10, 343-348, 1991									
A/Title: The hematopoietic and epithelial forms of CD44 are distinct polypeptides wi									
A/Reference number: S13530; MUID:91122041; PMID:1991450									
A/Accession: S13530									
A/Status: preliminary									
A/Molecule type: mRNA									
A/Residues: 1-493 <ST>									
A/Cross-references: EMBL:X5150; NID:929800; PIDN:CAA38951.1; PID:929801									
C/Keywords: transmembrane protein									
Query Match									
Best Local Similarity 99.3%; Score 2590; DB 2; Length 493;									
Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;									
QY	1	MDKFWHAAWGLCLVPLSLA	QIDNTICRAGVHVKNRYSISREADLCAPNSTL	60					
DB	1	MDKFWHAAWGLCLVPLSLA	QIDNTICRAGVHVKNRYSISREADLCAPNSTL	60					
QY	61	PTMAQMEKALSIGETCRVYGEHVVIPRIHNSICANNVTGYIITSTQYDYCFN	120						
DB	61	PTMAQMEKALSIGETCRVYGEHVVIPRIHNSICANNVTGYIITSTQYDYCFN	120						
QY	121	ASAPPEEDCTSVLPNAPFGPIITTVNDDGIRYVKGERTINPEDIYSPNPDVVSS	180						
DB	121	ASAPPEEDCTSVLPNAPFGPIITTVNDDGIRYVKGERTINPEDIYSPNPDVVSS	180						
QY	181	GSSERSSTSGCYIFFTFTVPIPDSDSPWITDTRIPATNDSHSTTLOPTANPT	240						
DB	181	GSSERSSTSGCYIFFTFTVPIPDSDSPWITDTRIPATNDSHSTTLOPTANPT	240						
QY	241	GLVBDLDRGTPLSTTTOQNSQSFSTHEGLEBKDPHTSTLSSNRNVTGGRDPNH	300						
DB	241	GLVBDLDRGTPLSTTTOQNSQSFSTHEGLEBKDPHTSTLSSNRNVTGGRDPNH	300						
QY	301	SEGSTHLEGGTSHYPTKESRTPIPTSAKTGSGFYTAATVGDNSNVRSLSGDDTF	360						
DB	301	SEGSTHLEGGTSHYPTKESRTPIPTSAKTGSGFYTAATVGDNSNVRSLSGDDTF	360						
QY	361	HPSGGSHRTTHSESDGSHSGQEGGANTTSGPIITPDIPEMLIILASLALIALIACIA	420						
DB	361	HPSGGSHRTTHSESDGSHSGQEGGANTTSGPIITPDIPEMLIILASLALIALIACIA	420						
QY	421	VNSRRRCQCKKLVINGNGAVEDRKSGGLNGEASKSQEWVHLVKNSSSTPPQFWTAD	480						
DB	421	VNSRRRCQCKKLVINGNGAVEDRKSGGLNGEASKSQEWVHLVKNSSSTPPQFWTAD	480						

QY 481 TRNLQNDMKIGV 493  
 Db 481 TRNLQNDMKIGV 493

## RESULT 2

137369  
 epican - human  
 C:Species: Homo sapiens (man)  
 C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 03-Aug-2001  
 C:Accession: 137369; S24631  
 R:Kugelman, L.C.; Ganguly, S.; Haggerty, J.G.; Weisman, S.M.; Milstone, L.M.  
 J: Invest. Dermatol. 99, 866-891, 1992  
 A:Title: The core protein of epican, a heparan sulfate proteoglycan on keratinocytes, is  
 A:Reference number: 137369; PMID:181868  
 A:Note: corrected and republished from J. Invest. Dermatol. 99, 381-385, 1992  
 A:Accession: 137369  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-699 <RES>  
 A:Cross-references: EMBL:X66733; NID:g31190; PID:g31191

Query Match 95.5%; Score 2490; DB 2; Length 699;  
 Best Local Similarity 70.4%; Pred. No. 2e-152;  
 Matches 492; Conservative 1; Mismatches 0; Indels 206; Gaps 1;

QY 1 MDKFWHAAAGLCVPLSLAQIDNITCRFAGVFHVENKGRYSISRTAADLCKAFNSTL 60  
 Db 1 MDKFWHAAAGLCVPLSLAQIDNITCRFAGVFHVENKGRYSISRTAADLCKAFNSTL 60  
 QY 61 PTMAQMERKALSIGFETCRGFFEGHVVPRIHPNSICANNVGVIILTSNTSOYDYTCFN 120  
 Db 61 PTMAQMERKALSIGFETCRGFFEGHVVPRIHPNSICANNVGVIILTSNTSOYDYTCFN 120  
 QY 61 PTMAQMERKALSIGFETCRGFFEGHVVPRIHPNSICANNVGVIILTSNTSOYDYTCFN 120  
 Db 61 PTMAQMERKALSIGFETCRGFFEGHVVPRIHPNSICANNVGVIILTSNTSOYDYTCFN 120  
 QY 121 ASAPBEDCTSTVDLPNAPFDGPTITTVNRDGTFRVQGEKRTPEDIYPSNPDDVSS 180  
 Db 121 ASAPBEDCTSTVDLPNAPFDGPTITTVNRDGTFRVQGEKRTPEDIYPSNPDDVSS 180  
 QY 121 ASAPBEDCTSTVDLPNAPFDGPTITTVNRDGTFRVQGEKRTPEDIYPSNPDDVSS 180  
 Db 121 ASAPBEDCTSTVDLPNAPFDGPTITTVNRDGTFRVQGEKRTPEDIYPSNPDDVSS 180  
 QY 181 GSSSRSSSTSGGYIYFTSTVHPRIPEDESFWITDSTRIPAT 222  
 Db 181 GSSSRSSSTSGGYIYFTSTVHPRIPEDESFWITDSTRIPAT 222  
 QY 181 GSSSRSSSTSGGYIYFTSTVHPRIPEDESFWITDSTRIPAT 222  
 Db 181 GSSSRSSSTSGGYIYFTSTVHPRIPEDESFWITDSTRIPAT 222  
 QY 223 223  
 Db 223 223  
 QY 241 DERDRLSFGSGIDDDDFISSTISTTPRAFDHTKQNDWTQNNPSHNPVLLQTTTR 300  
 Db 241 DERDRLSFGSGIDDDDFISSTISTTPRAFDHTKQNDWTQNNPSHNPVLLQTTTR 300  
 QY 223 223  
 Db 223 223  
 QY 301 MTDVDRNGTAYEGNWNPEAPPLIHHEHEEETPHSTSTIOATPSSTTEETATQKQW 360  
 Db 301 MTDVDRNGTAYEGNWNPEAPPLIHHEHEEETPHSTSTIOATPSSTTEETATQKQW 360  
 QY 223 223  
 Db 223 223  
 QY 361 FGNRWAVGRQTPKXDSHTTGTAAASHTSHPMQGRTPPSPEDSSWTFNPIISHMGR 420  
 Db 361 FGNRWAVGRQTPKXDSHTTGTAAASHTSHPMQGRTPPSPEDSSWTFNPIISHMGR 420  
 QY 223 223  
 Db 223 223  
 QY 421 GHQGRBMDSDSHSTTLQPTANPMTGLVEDLDRTPGLSMTTQOSNSQSFSTSHGLEED 480  
 Db 421 GHQGRBMDSDSHSTTLQPTANPMTGLVEDLDRTPGLSMTTQOSNSQSFSTSHGLEED 480  
 QY 275 KQHPTSTLTSNRNDVYGGRRDPNHSSTTLLEGYSHYPTKESRFFIYTSKATGS FGVTAVYVDSNNVR 334  
 Db 275 KQHPTSTLTSNRNDVYGGRRDPNHSSTTLLEGYSHYPTKESRFFIYTSKATGS FGVTAVYVDSNNVR 334  
 QY 481 KQHPTSTLTSNRNDVYGGRRDPNHSSTTLLEGYSHYPTKESRFFIYTSKATGS FGVTAVYVDSNNVR 540  
 Db 481 KQHPTSTLTSNRNDVYGGRRDPNHSSTTLLEGYSHYPTKESRFFIYTSKATGS FGVTAVYVDSNNVR 540  
 QY 335 FGVTAVYVDSNNVRISLGDQDTFHPGSGSHHTHGSBDGSHSGOGGANTTSGPIR 394  
 Db 335 FGVTAVYVDSNNVRISLGDQDTFHPGSGSHHTHGSBDGSHSGOGGANTTSGPIR 394  
 QY 541 FGVTAVYVDSNNVRISLGDQDTFHPGSGSHHTHGSBDGSHSGOGGANTTSGPIR 600  
 Db 541 FGVTAVYVDSNNVRISLGDQDTFHPGSGSHHTHGSBDGSHSGOGGANTTSGPIR 600  
 QY 395 TQOIPRWLIIILASLALAILAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEA 454  
 Db 395 TQOIPRWLIIILASLALAILAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEA 454  
 QY 601 TQOIPRWLIIILASLALAILAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEA 660  
 Db 601 TQOIPRWLIIILASLALAILAVCTAVNRRRCQKKLVINSNGAVEDRKPGLNGEA 660  
 QY 455 SKSQEMVHLVKNKSESTPDQFMTADETRNLQNDMKIGV 493  
 Db 455 SKSQEMVHLVKNKSESTPDQFMTADETRNLQNDMKIGV 493

Db 661 SKSQEMVHLVKNKSESTPDQFMTADETRNLQNDMKIGV 699

## RESULT 3

A47195  
 lymphocyte homing receptor isoform CD44 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Nov-1994  
 C:Accession: A47195  
 R:Scrutton, G.R.; Bell, M.V.; Jackson, D.G.; Cornelius, F.B.; Gerth, U.; Bell, J.I.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 12160-12164, 1992  
 A:Title: Genomic structure of DNA encoding the lymphocyte homing receptor CD44 revea  
 A:Reference number: A47195; MUID:93101687; PMID:1465456  
 A:Accession: A47195  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-742 <SCR>  
 A:Note: sequence inconsistent with the nucleotide translation  
 A:Note: sequence extracted from NCBI Backbone (NCBIN:120731, NCBIN:120737, NCBIN:12  
 A:NCBIN:120764, NCBIN:120766, NCBIN:120770, NCBIN:120772, NCBIN:120774, NCBIN:12077

Query Match 94.7%; Score 2468.5; DB 2; Length 742;  
 Best Local Similarity 66.3%; Pred. No. 5.1e-151;  
 Matches 492; Conservative 1; Mismatches 0; Indels 249; Gaps 1;

QY 1 MDKFWHAAAGLCVPLSLAQIDNITCRFAGVFHVENKGRYSISRTAADLCKAFNSTL 60  
 Db 1 MDKFWHAAAGLCVPLSLAQIDNITCRFAGVFHVENKGRYSISRTAADLCKAFNSTL 60  
 QY 61 PTMAQMERKALSIGFETCRGFFEGHVVPRIHPNSICANNVGVIILTSNTSOYDYTCFN 120  
 Db 61 PTMAQMERKALSIGFETCRGFFEGHVVPRIHPNSICANNVGVIILTSNTSOYDYTCFN 120  
 QY 61 PTMAQMERKALSIGFETCRGFFEGHVVPRIHPNSICANNVGVIILTSNTSOYDYTCFN 120  
 Db 61 PTMAQMERKALSIGFETCRGFFEGHVVPRIHPNSICANNVGVIILTSNTSOYDYTCFN 120  
 QY 121 ASAPBEDCTSTVDLPNAPFDGPTITTVNRDGTFRVQGEKRTPEDIYPSNPDDVSS 180  
 Db 121 ASAPBEDCTSTVDLPNAPFDGPTITTVNRDGTFRVQGEKRTPEDIYPSNPDDVSS 180  
 QY 121 ASAPBEDCTSTVDLPNAPFDGPTITTVNRDGTFRVQGEKRTPEDIYPSNPDDVSS 180  
 Db 121 ASAPBEDCTSTVDLPNAPFDGPTITTVNRDGTFRVQGEKRTPEDIYPSNPDDVSS 180  
 QY 181 GSSSRSSSTSGGYIYFTSTVHPRIPEDESFWITDSTRIPAT 222  
 Db 181 GSSSRSSSTSGGYIYFTSTVHPRIPEDESFWITDSTRIPAT 222  
 QY 181 GSSSRSSSTSGGYIYFTSTVHPRIPEDESFWITDSTRIPAT 222  
 Db 181 GSSSRSSSTSGGYIYFTSTVHPRIPEDESFWITDSTRIPAT 222  
 QY 223 223  
 Db 223 223  
 QY 241 TWDPSWFLPESRNHLHTTTSNTISAGWEPNEEDTOMAGERDRLSFGSGIDDD 300  
 Db 241 TWDPSWFLPESRNHLHTTTSNTISAGWEPNEEDTOMAGERDRLSFGSGIDDD 300  
 QY 223 223  
 Db 223 223  
 QY 301 EDPISSTITTRADHTKQNDWTQNNPSHNPVLLQTTTRMTDVRNGTAYEGNWN 360  
 Db 301 EDPISSTITTRADHTKQNDWTQNNPSHNPVLLQTTTRMTDVRNGTAYEGNWN 360  
 QY 223 223  
 Db 223 223  
 QY 361 PEAPPLIHHEHEEETPHSTSTIOATPSSTTEETATQKQWFGNRMHEGYRQTPRDS 420  
 Db 361 PEAPPLIHHEHEEETPHSTSTIOATPSSTTEETATQKQWFGNRMHEGYRQTPRDS 420  
 QY 223 223  
 Db 223 223  
 QY 421 HSTTGTAASANTSHPMQGRTPPSPEDSSWTFNPIISHMGRGHQGRBMDSDSHST 480  
 Db 421 HSTTGTAASANTSHPMQGRTPPSPEDSSWTFNPIISHMGRGHQGRBMDSDSHST 480  
 QY 232 LQPTANPMTGLVEDLDRTPGLSMTTQOSNSQSFSTSHGLEEDKQHPPTSTLTSNRNDV 291  
 Db 232 LQPTANPMTGLVEDLDRTPGLSMTTQOSNSQSFSTSHGLEEDKQHPPTSTLTSNRNDV 291  
 QY 481 LQPTANPMTGLVEDLDRTPGLSMTTQOSNSQSFSTSHGLEEDKQHPPTSTLTSNRNDV 540  
 Db 481 LQPTANPMTGLVEDLDRTPGLSMTTQOSNSQSFSTSHGLEEDKQHPPTSTLTSNRNDV 540  
 QY 292 TQGRDPNHSSTTLLEGYSHYPTKESRFFIYTSKATGS FGVTAVYVDSNNVR 351  
 Db 292 TQGRDPNHSSTTLLEGYSHYPTKESRFFIYTSKATGS FGVTAVYVDSNNVR 351  
 QY 541 TQGRDPNHSSTTLLEGYSHYPTKESRFFIYTSKATGS FGVTAVYVDSNNVR 600  
 Db 541 TQGRDPNHSSTTLLEGYSHYPTKESRFFIYTSKATGS FGVTAVYVDSNNVR 600  
 QY 352 SLGDDQDTFHPGSGSHHTHGSBDGSHSGOGGANTTSGPIRTPQIPRWLIIILASLAL 411  
 Db 352 SLGDDQDTFHPGSGSHHTHGSBDGSHSGOGGANTTSGPIRTPQIPRWLIIILASLAL 411  
 QY 601 SLGDDQDTFHPGSGSHHTHGSBDGSHSGOGGANTTSGPIRTPQIPRWLIIILASLAL 660  
 Db 601 SLGDDQDTFHPGSGSHHTHGSBDGSHSGOGGANTTSGPIRTPQIPRWLIIILASLAL 660  
 QY 412 ALLAVCIANVRRRCQKKLVINSNGAVEDRKPGLNGEASKSQEMVHLVKNKSEST 471  
 Db 412 ALLAVCIANVRRRCQKKLVINSNGAVEDRKPGLNGEASKSQEMVHLVKNKSEST 471  
 QY 661 ALLAVCIANVRRRCQKKLVINSNGAVEDRKPGLNGEASKSQEMVHLVKNKSEST 720  
 Db 661 ALLAVCIANVRRRCQKKLVINSNGAVEDRKPGLNGEASKSQEMVHLVKNKSEST 720

QY 472 PDQFMTADERTNLCQVDMKIGV 493  
 Db 721 PDQFMTADERTNLCQVDMKIGV 742

## RESULT 4

JH0518  
 Lymphocyte homing receptor CD44, splice form CD44R1 - human  
 N:Alternate names: cell adhesion molecule core protein CD44E, keratinocyte, cell surface  
 N:Contains: Lymphocyte homing receptor CD44, splice form CD44R1, Lymphocyte homing recep  
 C:Species: Homo sapiens (man)  
 C>Date: 30-Jun-1992 #sequence revision 30-Jun-1992 #text change 18-Aug-2000  
 C:Accession: JH0518; JH0519; PH0859; A39209; A42402; C42402; A53029; S16147  
 R:Dougherty, G.J.; Lansdorf, P.M.; Cooper, D.L.; Humphries, R.K.  
 J. Exp. Med. 174, 1-5, 1991  
 A:Title: Molecular cloning of CD44R1 and CD44R2, two novel isoforms of the human CD44 ly  
 A:Reference number: JH0518; MUID:91277598; PMID:2056274  
 A:Accession: JH0518  
 A:Molecule type: mRNA  
 A:Residues: 1-426 <DOU>  
 A:Experimental source: Lymphocytes, cell line KG1a  
 A:Accession: JH0519  
 A:Molecule type: mRNA  
 A:Residues: 1-223,288-426 <DO2>  
 A:Experimental source: Lymphocyte, cell line KG1a  
 R:Cooper, D.L.; Dougherty, G.; Hahn, H.J.; Jackson, S.; Baptist, E.W.; Byers, J.; Datta,  
 Biochem. Biophys. Res. Commun. 182, 569-578, 1992  
 A:Title: The complex CD44 transcripional unit: alternative splicing of three internal e  
 A:Reference number: PH0859; MUID:92134271; PMID:1734871  
 A:Accession: PH0859  
 A:Molecule type: DNA  
 A:Residues: 223-357 <COO>  
 R:Brown, T.A.; Bouchard, T.; St. John, T.; Wayner, E.; Carter, W.G.  
 J. Cell Biol. 113, 207-221, 1991  
 A:Title: Human Keratinocytes express a new CD44 core protein (CD44E) as a heparan-sulfat  
 A:Reference number: A39209; MUID:91177958; PMID:2007624  
 A:Accession: A39209  
 A:Molecule type: mRNA  
 A:Residues: 184-376 <BRO>  
 A:Cross-references: GB:X55938; NID:929802; PIDN:CAA39404.1; PID:9390047  
 R:Jackson, D.G.; Buckley, J.; Bell, J.T.  
 J. Biol. Chem. 267, 4732-4739, 1992  
 A:Title: Multiple variants of the human lymphocyte homing receptor CD44 generated by ins  
 A:Reference number: A42402; MUID:92165834; PMID:1537855  
 A:Accession: A42402  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 217-223,288-359 <JNC>  
 A:Note: sequence extracted from NCBI backbone (NCBI:N:83964, NCBI:P:83965)  
 A:Accession: C42402  
 A:Note: variant B  
 A:Accession: C42402  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 217-370 <JAZ>  
 A:Note: sequence extracted from NCBI backbone (NCBI:N:83968, NCBI:P:83969)  
 A:Note: variant D  
 R:Shepley, M.P.; Racanelli, V.R.  
 J. Virol. 68, 1301-1308, 1994  
 A:Title: A monoclonal antibody that blocks poliovirus attachment recognizes the lymphocy  
 A:Reference number: A53029; MUID:94149816; PMID:7508992  
 A:Accession: A53029  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 67-76, 'X', 78-89 <SHB>  
 C:Genetics:  
 A:Gene: GDB:CD44; MDU2; MDU3; MI  
 A:Cross-references: GDB:120739; OMIM:107269  
 A:Map position: 11pter-11p13  
 A:Introns: 35/1; 65/1; 133/1  
 C:Superfamily: human cell adhesion protein CD44  
 C:Keywords: alternative splicing; cell adhesion; chondroitin sulfate proteoglycan; glyco  
 F:1-426/Product: Lymphocyte homing receptor CD44, splice form CD44R1 #status predicted <  
 F:1-223,288-426/Product: Lymphocyte homing receptor CD44, splice form CD44R2 #status pre

F:299/Binding site: carbohydrate (asn) (covalent) #status predicted  
 F:354/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 86.7%; Score 2261; DB 2; Length 426;

Best Local Similarity 100.0%; Pred. No. 5,4e-138; Indels 0; Gaps 0;

Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDKFWMAHAWGLCLVPLSLAQIDNITCRFAGVHVKNGRYSISRTAADLCRAFNSTL 60  
 Db 1 MDKFWMAHAWGLCLVPLSLAQIDNITCRFAGVHVKNGRYSISRTAADLCRAFNSTL 60  
 QY 61 PTMAQMERKALSIGETCRGYSIBGVVPIRHPNSICANNNGYIILTSNTSOYDTYCFN 120  
 Db 61 PTMAQMERKALSIGETCRGYSIBGVVPIRHPNSICANNNGYIILTSNTSOYDTYCFN 120  
 QY 121 ASAPPEEDCTSVTLDPNAPDPIITITVNRDGTREYVKGEXRTNPEDIYSPNPTDDVSS 180  
 Db 121 ASAPPEEDCTSVTLDPNAPDPIITITVNRDGTREYVKGEXRTNPEDIYSPNPTDDVSS 180  
 QY 181 GSSSRFSTSGCYIFFTSTYHPIPDSDSPMITTSTDRIPATMNDSSHSTLLOTPANT 240  
 Db 181 GSSSRFSTSGCYIFFTSTYHPIPDSDSPMITTSTDRIPATMNDSSHSTLLOTPANT 240  
 QY 241 GLVEDLDRTGPLSMWTQSNQSFSTSHGLEDKDPHTTSTLTSNRNDVYQGRDPNH 300  
 Db 241 GLVEDLDRTGPLSMWTQSNQSFSTSHGLEDKDPHTTSTLTSNRNDVYQGRDPNH 300  
 QY 301 SBGSTTLLEGTTSHYPTKESRTPIPTVSANTGSPGTAVTVGSSNVNRSLSGDQTF 360  
 Db 301 SBGSTTLLEGTTSHYPTKESRTPIPTVSANTGSPGTAVTVGSSNVNRSLSGDQTF 360  
 QY 361 HPSGGSHTTHGSESDGHSQEGGANTTSPITTPQIPFWLILASLALALILAVCIA 420  
 Db 361 HPSGGSHTTHGSESDGHSQEGGANTTSPITTPQIPFWLILASLALALILAVCIA 420  
 QY 421 VNRR 425  
 Db 421 VNRR 425

RESULT 5  
 17371  
 CD44R5 - human  
 C:Species: Homo sapiens (man)  
 C>Date: 02-Aug-1996 #sequence revision 02-Aug-1996 #text change 21-Jul-2000  
 C:Accession: I77371  
 M.Oi. Carcinog. 7, 212-220, 1993  
 A:Title: Novel variants of CD44 arising from alternative splicing: changes in the C  
 A:Reference number: I57483; MUID:93356912; PMID:8352881  
 A:Accession: I77371  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-395 <RES>  
 A:Cross-references: GB:S66400; NID:9435697; PIDN:AAH27919.1; PID:9435700  
 C:Genetics:  
 A:Gene: GDB:CD44  
 A:Cross-references: GDB:120739; OMIM:107269  
 A:Map position: 11pter-11p13  
 A:Introns: 257/1  
 C:Superfamily: human cell adhesion protein CD44

Query Match 78.1%; Score 2036; DB 2; Length 395;  
 Best Local Similarity 79.9%; Pred. No. 1,4e-123; Indels 98; Gaps 1;  
 Matches 394; Conservative 1; Mismatches 0;

QY 1 MDKFWMAHAWGLCLVPLSLAQIDNITCRFAGVHVKNGRYSISRTAADLCRAFNSTL 60  
 Db 1 MDKFWMAHAWGLCLVPLSLAQIDNITCRFAGVHVKNGRYSISRTAADLCRAFNSTL 60  
 QY 61 PTMAQMERKALSIGETCRGYSIBGVVPIRHPNSICANNNGYIILTSNTSOYDTYCFN 120  
 Db 61 PTMAQMERKALSIGETCRGYSIBGVVPIRHPNSICANNNGYIILTSNTSOYDTYCFN 120

QY	121	ASAPPEECSTVTDLPNAPFDGPITTTIVNRGSTRVVGKGEYRNPEDYIPENPDDVSS	180
Dd	121	ASAPPEECSTVTDLPNAPFDGPITTTIVNRGSTRVVGKGEYRNPEDYIPENPDDVSS	180
QY	181	GSSSRSSSTGCGYIFETSTVHPIDEDSPWITDSTRIPATNNDSSHITLOPTANENT	240
Dd	181	GSSSRSSSTGCGYIFETSTVHPIDEDSPWITDSTRIPATNNDSSHITLOPTANENT	240
QY	241	GLVEBLDPTGPLSMITQOONSQSFTSHGLEGEDKDHFTTSTLSSNRNDVTGGRDPNH	300
Dd	241	GLVEBLDPTGPLSMITR-----	257
QY	301	SEGSTTLEGYTSHYPHTKESRTFIPVTSAXTSGPVTAIVVGDSNSVNRSLGODOPT	360
Dd	258	-----DQDTF	262
QY	361	HPSGGSHHTHGSSEDDSHSGSQEGGANTTSGPIRTPQPEWIIILASILALAILAVCIA	420
Dd	263	HPSGGSHHTHGSSEDDSHSGSQEGGANTTSGPIRTPQPEWIIILASILALAILAVCIA	322
QY	421	VNSRRRCQKKLVINSNGAVEDRKRPGLNGEASKSQSEMHVLYNKESSETPDQPMTAD	480
Dd	323	VNSRRRCQKKLVINSNGAVEDRKRPGLNGEASKSQSEMHVLYNKESSETPDQPMTAD	382
QY	481	TRNLQNVDMKIGV	493
Dd	383	TRNLQNVDMKIGV	395

## RESULT 6

cell adhesion molecule CD44 - human

C:Date: 23-Nov-1991 #sequence\_revision 23-Nov-1991 #text\_change 01-Dec-2000

C:Accession: UH0417; A32376; G02251; A32377

R:Ham, H.U.; Isola, N.; Cooper, D.L.

Biochem. Biophys. Res. Commun. 178, 1127-1134, 1991

A:Title: The multispecific cell adhesion molecule CD44 is represented in reticulocyte c

A:Reference number: UH0417; MUID:91337049; PMID:1840487

A:Accession: UH0417

A:Molecule type: mRNA

A:Residues: 1-361 <HAR>

A:CROSS-references: GB:M59040; NID:G180129; PIDN:AAA5150.1; PID:G180130

A:Experimental source: reticulocyte

A:Note: the authors translated the codon ATG for residues 63, 66 and 239 as Trp and GGA

R:Stamenkovic, I.; Amiot, M.; Pesando, J.M.; Seed, B.

Cell 56, 1057-1062, 1989

A:Title: A lymphocyte molecule implicated in lymph node homing is a member of the cartil

A:Reference number: A32376; MUID:89168434; PMID:2466575

A:Accession: A32376

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-238, 'E', 240-361 <STA>

A:CROSS-references: GB:M4915; NID:G180196; PIDN:AAA35674.1; PID:G180197

R:Bosch, P.P.; Stevens, J.W.; Buckwalter, J.A.; Mitura, R.J.

submitted to the EMBL Data Library, November 1995

A:Reference number: H00921

A:Accession: G02251

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-25, 'M', 27-108, 'S', 110-361 <BOS>

A:CROSS-references: EMBL:U40373; NID:G1101785; PID:G1101786

R:Goldstein, L.A.; Zhou, D.F.H.; Picker, L.U.; Minny, C.N.; Bargatze, R.F.; Ding, J.F.;

Cell 56, 1063-1072, 1989

A:Title: A human lymphocyte homing receptor, the herpes antigen, is related to cartilage

A:Reference number: A32377; MUID:89168435; PMID:2466576

A:Accession: A32377

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-108, 'S', 110-293, 'S', <GOL>

A:CROSS-references: GB:M25078; NID:G186660; PIDN:AAA36138.1; PID:G186661

C:Superfamily: human cell adhesion protein CD44

```
C:\Keywords: alternative splicing; cell adhesion; surface antigen; transmembrane protein;
F1269-285/Domain: transmembrane #status predicted <TM>
Query Match 70.4%; Score 1835; DB 2; Length 361;
Best Local Similarity 72.8%; Pred. No. 1e-110;
Matches 359; Conservative 0; Mismatches 2; Indels 132; Gaps 1;
```

```

0Y      1 MDFFNNHAAWGLCLVPLSLAQIDILNITRFGAVFVFNXGYSISRTPEADLCGFANSTL 60
Db      1 MDFFNNHAAWGLCLVPLSLAQIDILNITRFGAVFVFNXGYSISRTPEADLCGFANSTL 60
0Y      61 PTMAQNEKSLSTIGFEFCRGTGEGHVVIPIRIHNSICAAANTGVYIILTSNISOVDTCFN 120
Db      61 PTMAQNEKSLSTIGFEFCRGTGEGHVVIPIRIHNSICAAANTGVYIILTSNISOVDTCFN 120
0Y      121 ASAPPEEDCTSYDLENAPFDGPIITLTVNRDQTRVQGEKRYNPEDIPENPTDDVSS 180
Db      121 ASAPPEEDCTSYDLENAPFDGPIITLTVNRDQTRVQGEKRYNPEDIPENPTDDVSS 180
0Y      181 GSSSEFSSSTSGGIPTFTBSTVHPIDEDSPMTTDSIIPATNMDSSHTLQETAPNT 240
Db      181 GSSSEFSSSTSGGIPTFTBSTVHPIDEDSPMTTDSIIPATNMDSSHTLQETAPNT 223
0Y      241 GLVEDIDRGPFLMTQQSNQSQSFSTSHGLIEDKHPTSTLTSNENVDVTGGRDPNH 300
Db      241 ----- 223
0Y      301 SEGSSTLLBGYSHPHTKESRTPIPVTSANTGFCVTATVGDNSNVNRSLSGDDTF 360
Db      301 -----DDDTF 228
0Y      361 HPSCGSHTHGSESDSHSGSOEGAGNTTSGPIRTPQIPEWLIILASLLALILAVCIA 420
Db      229 HPSCGSHTHGSESDSHSGSOEGAGNTTSGPIRTPQIPEWLIILASLLALILAVCIA 288
0Y      421 VNSRRRCQGRKKLVINSNGAVEDEKPSGLNGEASKSOEMVILYNKSSSETPDQFMTADE 480
Db      289 VNSRRRCQGRKKLVINSNGAVEDEKPSGLNGEASKSOEMVILYNKSSSETPDQFMTADE 348
0Y      481 TRNLQNVDMKIGV 493
Db      349 TRNLQNVDMKIGV 361

```

## RESULT 7

Lymphocyte adhesion receptor precursor - baboon

C1Species: Papio sp. (baboon)

C1Date: 18-Apr-1989 #sequence\_revision 18-Apr-1989 #text\_change 21-Jul-2000

C1Accession: A33935; A30901

C1Riderera, R.L.; Catter, W.G.; Nottenburg, C.; Wayner, E.A.; Gallatin, W.M.; St. John

Proc. Natl. Acad. Sci. U.S.A. 86; 4659-4663, 1989

A1Title: Isolation and DNA sequence of a cDNA clone encoding a lymphocyte adhesion

A1Reference number: A33935; PMID:89282830; PMID:2471974

A1Accession: A33935

A1Status: preliminary

A1Molecule type: mRNA

A1Residues: 1-362 <IDZ>

A1Cross-references: GB:M22452; NID:G176576; PID:G176577

C1Comment: This protein was isolated from the herpes papio induced B cell lymphoma.

C1Genetics:

A1Gene: CD44; ECCR11; Hermes-1 antigen

C1Superfamily: human cell adhesion protein CD44

C1Keywords: cytoskeleton; extracellular matrix; glycoprotein; lipoprotein; surface

F1-20/Domain: signal sequence #status predicted <SIG>

F121-362/Product: lymphocyte adhesion receptor #status predicted <MAT>

F121-270/Domain: extracellular #status predicted <EXT>

F121-290/Domain: transmembrane #status predicted <TM>

F1291-362/Domain: intracellular #status predicted <CYT>

F125-57,100,110,120,256/Binding site: carbohydrate (Asn) (covalent) #status predicted

F1296/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 65.1%; Score 1698.5; DB 2; Length 362;

Best Local Similarity 68.2%; Pred. No. 6,1e-102; Indels 133; Gaps 2;  
Matches 337; Conservative 7; Mismatches 17;

QY 1 MDKFWMAAMGLCLVPLSLAQIDLNITCRFAGVFNHKNRGYSISRTAADLCKAFNSTL 60  
DB 1 MDKFWMAAMGLCLVPLSLAQIDLNITCRFAGVFNHKNRGYSISRTAADLCKAFNSTL 60  
QY 61 PTMAQMEKALSIGETCRGFIHGVIPRIHPNSICANNNGVYLLTSTSGYDTCRN 120  
DB 61 PTMAQMEKALSIGETCRGFIHGVIPRIHPNSICANNNGVYLLTSTSGYDTCRN 120  
QY 121 ASAPPEEDCTSVTLDPNAPFDGPITITIVNRDGRYVQKGEYRNPEDIPSPNPTDDVSS 180  
DB 121 ASAPPEEDCTSVTLDPNAPFDGPITITIVNRDGRYVQKGEYRNPEDIPSPNPTDDVSS 180  
QY 121 ASAPPEEDCTSVTLDPNAPFDGPITITIVNRDGRYVQKGEYRNPEDIPSPNPTDDVSS 180  
DB 121 ASAPPEEDCTSVTLDPNAPFDGPITITIVNRDGRYVQKGEYRNPEDIPSPNPTDDVSS 180  
QY 181 GSSSRSSSTGGYIFFT-FSTVHPIDPDESPWITDSDRIIPATNMSSHTLQPTANEN 239  
DB 181 GSSSRSSSTGGYIFFT-FSTVHPIDPDESPWITDSDRIIPATNMSSHTLQPTANEN 239  
QY 240 TGLVEDLDRGTPLSMITQGSNSQSFSTSHGLEEDKDPPTSTLTSNRNDVYGRDEN 299  
DB 240 TGLVEDLDRGTPLSMITQGSNSQSFSTSHGLEEDKDPPTSTLTSNRNDVYGRDEN 299  
QY 300 HSEGSTLLSGYISHYPTKESRTFIPVTSKATGSGCVTAVTGDSNSVNRSLSGDDOT 359  
DB 300 HSEGSTLLSGYISHYPTKESRTFIPVTSKATGSGCVTAVTGDSNSVNRSLSGDDOT 359  
QY 360 FHPEGSHTHGSDSHSGOEGGANTTSGPIRTPQIPEWLLIILASLALAILAVCI 419  
DB 229 FDBPGSHTHGSDSHSGOEGGANTTSGPIRTPQIPEWLLIILASLALAILAVCI 288  
QY 420 AVNSRRRCGQKKKLVINGAVDRKPSGLNGEASKSQEMVHLVKNSEETPDQFMTAD 479  
DB 289 AVNSRRRCGQKKKLVINGAVDRKPSGLNGEASKSQEMVHLVKNSEETPDQFMTAD 348  
QY 480 ETRRLQNVDMKIGV 493  
DB 349 ETRRLQNVDMKIGV 362

# RESULT 8

224240  
Lymphocyte surface antigen CD44 precursor - horse

C:Species: Equus caballus (domestic horse)

C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 21-Jul-2000

C/Accession: I46245; S24240

R/Tavernor, A.S.; Deyerson, E.V.; Coadwell, W.J.; Lunn, D.P.; Zhang, C.; Davis, W.; Butc

Immunogenetics 37, 474-477, 1993

A>Title: Molecular cloning of equine CD44 cDNA by a COS cell expression system.

A/Reference number: I46245; MUID:93170897; PMID:8436424

A/Accession: I46245

A>Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: mRNA

A/Residues: 1-359 <T2>

A/Cross-references: EMBL:X66862; NID:91059; PIDN:CAA47331.1; PID:91060

C/Superfamily: human cell adhesion protein CD44

C/Keywords: surface antigen; transmembrane protein

Query Match 59.3%; Score 1547; DB 2; Length 359;

Best Local Similarity 63.0%; Pred. No. 3.3e-92; Indels 138; Gaps 5;

Matches 312; Conservative 20; Mismatches 25;

QY 1 MDKFWMAAMGLCLVPLSLAQIDLNITCRFAGVFNHKNRGYSISRTAADLCKAFNSTL 60  
DB 1 MDKFWMAAMGLCLVPLSLAQIDLNITCRFAGVFNHKNRGYSISRTAADLCKAFNSTL 60  
QY 61 PTMAQMEKALSIGETCRGFIHGVIPRIHPNSICANNNGVYLLTSTSGYDTCRN 120  
DB 61 PTMAQMEKALSIGETCRGFIHGVIPRIHPNSICANNNGVYLLTSTSGYDTCRN 120  
QY 121 ASAPPEEDCTSVTLDPNAPFDGPITITIVNRDGRYVQKGEYRNPEDIPSPNPTDDVSS 180  
DB 121 ASAPPEEDCTSVTLDPNAPFDGPITITIVNRDGRYVQKGEYRNPEDIPSPNPTDDVSS 180

# RESULT 9

545305  
CD44 antigen precursor - dog (fragment)

C:Species: Canis lupus familiaris (dog)

C/Date: 20-Oct-1994 #sequence\_revision 21-Jul-1995 #text\_change 21-Jul-2000

C/Accession: S45305

R/Milde, K.F.; Alejandro, R.; Mintz, D.H.; Pastori, R.L.

Biochem. Biophys. Acta 1218, 112-114, 1994

A>Title: Molecular cloning of the canine CD44 antigen cDNA.

A/Reference number: S45305; MUID:94250687; PMID:7514890

A/Accession: S45305

A>Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-351 <M1>

A/Cross-references: EMBL:Z27115; NID:9473226; PIDN:CAA61630.1; PID:9473227

C/Superfamily: human cell adhesion protein CD44

Query Match 56.4%; Score 1470; DB 2; Length 351;

Best Local Similarity 61.7%; Pred. No. 2.9e-87; Indels 132; Gaps 3;

Matches 297; Conservative 21; Mismatches 31;

QY 9 AMGLCLVPLSLAQIDLNITCRFAGVFNHKNRGYSISRTAADLCKAFNSTLPTMAQMEK 68  
DB 2 AMGLCLVPLSLAQIDLNITCRFAGVFNHKNRGYSISRTAADLCKAFNSTLPTMAQMEK 61  
QY 69 ALSTGFETCRGFIHGVIPRIHPNSICANNNGVYLLTSTSGYDTCRNASAPPEED 128  
DB 62 ALSTGFETCRGFIHGVIPRIHPNSICANNNGVYLLTSTSGYDTCRNASAPPEED 121  
QY 129 CTSVTLDPNAPFDGPITITIVNRDGRYVQKGEYRNPEDIPSPNPTDDVSSGSSSRSS 188  
DB 122 CTSVTLDPNAPFDGPITITIVNRDGRYVQKGEYRNPEDIPSPNPTDDVSSGSSSRSS 181  
QY 189 TSGYIFFT-FSTVHPIDPDESPWITDSDRIIPATNMSSHTLQPTANNTGLVELD 247  
DB 182 TSGYIFFT-FSTVHPIDPDESPWITDSDRIIPATNMSSHTLQPTANNTGLVELD 216  
QY 248 RTGPLSMITQGSNSQSFSTSHGLEEDKDPPTSTLTSNRNDVYGRDENHSGSTTL 307  
DB 217 RTGPLSMITQGSNSQSFSTSHGLEEDKDPPTSTLTSNRNDVYGRDENHSGSTTL 219  
QY 308 LEGYTSHYPTKESRTFIPVTSKATGSGCVTAVTGDSNSVNRSLSGDDOTFHPSGSH 367  
DB 220 LEGYTSHYPTKESRTFIPVTSKATGSGCVTAVTGDSNSVNRSLSGDDOTFHPSGSH 230  
QY 368 TTHGSDSHSGOEGGANTTSGPIRTPQIPEWLLIILASLALAILAVCIAVNSRRRC 427  
DB 368 TTHGSDSHSGOEGGANTTSGPIRTPQIPEWLLIILASLALAILAVCIAVNSRRRC 427

Db 231 TTHSGSAGHSSGQEGANTTSGPMKPGQIPMLIIILASLALILAVCIANVSRRC 290

Qy 428 GQKKVYINSGNVAVEDRKPSGLNGEASKSQEMVHLVNTKSSSTPPOFMTADETRLQNV 487

Db 291 GQKKVYINSGNVAVEDRKPSGLNGEASKSQEMVHLVNTKSSSTPPOFMTADETRLQNV 350

Qy 488 D 488

Db 351 D 351

## RESULT 10

A53286  
cell-surface glycoprotein CD44 precursor - bovine  
N:Alternate names: CD44 protein  
C:Species: Bos primigenius taurus (cattle)  
C/Date: 02-May-1994 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
C/Accession: A53286; S22123  
R:Boeworth, B.T.; St John, T.; Gallatin, W.M.; Harp, J.A.  
Mol. Immunol. 28, 1131-1135, 1991  
A:Title: Sequence of the bovine CD44 cDNA: comparison with human and mouse sequences.  
A:Reference number: A53286; MUID:92017904; PMID:1922105  
A:Accession: A53286  
A:Molecule type: mRNA  
A:Residues: 1-366 <BOS>  
A:Cross-References: EMBL:X62881; NID:g186; PIDN:CA44675.1; PID:g187  
A:Note: Sequence extracted from NCBI backbone (NCBIN:63418, NCBI:P:63419)  
C:Superfamily: human cell adhesion protein CD44  
C:Keywords: cell adhesion; glycoprotein; transmembrane protein  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-366/Product: cell-surface glycoprotein CD44 #status predicted <MAT>  
F:274-294/Domain: transmembrane #status predicted <TM>  
F:25,57,100,110,120,222,260/Binding site: carbohydrate (asn) (covalent) #status predicted

## Query Match

Best Local Similarity 56.3%; Score 1469.5; DB 2; Length 366;  
Matches 305; Conservative 22; Mismatches 32; Indels 141; Gaps 8;

Qy 1 MDKFWMAHMGCLVPLSLA--QIDNITCRFAGVFHVEKNGRYSISRTAADLCKAFNSTL 60

Db 1 MDKFWMAHMGCLVPLSLA--QIDNITCRFAGVFHVEKNGRYSISRTAADLCKAFNSTL 60

Qy 61 PTMAQMEKALSIGFETCRGFIEGHVVIRIHPNSICANNNGVYIITNTSGYDTYCEN 120

Db 61 PTMAQMEKALSIGFETCRGFIEGHVVIRIHPNSICANNNGVYIITNTSGYDTYCEN 120

Qy 121 ASAPBEDCTSVTDLPLNAFDGPTITIVNRDGRVYVKGERTNEDIPSPNPTDDV 175

Db 121 ASAPBEDCTSVTDLPLNAFDGPTITIVNRDGRVYVKGERTNEDIPSPNPTDDV 180

Qy 176 DVVSSGSSERSSTISGCI-FYFT-FSTVHPIDEDSPWITDSTRIPATNMDSHSTTLQ 233

Db 176 DVVSSGSSERSSTISGCI-FYFT-FSTVHPIDEDSPWITDSTRIPATNMDSHSTTLQ 233

Qy 181 DEMSSGSPSER-STSGGYSIFHTLPTVHFRPR-RPW-----SORAENT----- 223

Db 181 DEMSSGSPSER-STSGGYSIFHTLPTVHFRPR-RPW-----SORAENT----- 223

Qy 224 PTANPTGLVEDLDRGTPLSMITQOONSQSFSTHEGLEEDKHPTSTILSNRNDVYG 233

Db 224 PTANPTGLVEDLDRGTPLSMITQOONSQSFSTHEGLEEDKHPTSTILSNRNDVYG 233

Qy 224 -----SSTRDYGSND----- 234

Db 224 -----SSTRDYGSND----- 234

Qy 294 GRDPNHSSEGSTLLLEGYTSHPHTKESRTPIPVTSAKTGSFGVAVTVGDSNVNRS 353

Db 294 GRDPNHSSEGSTLLLEGYTSHPHTKESRTPIPVTSAKTGSFGVAVTVGDSNVNRS 353

Qy 235 ----- 234

Db 235 ----- 234

Qy 334 SGDDTFHPSGSGSRTHSGSDGSHSGQEGANTTSGPIRTPOIPMLIIILASLALIL 413

Db 334 SGDDTFHPSGSGSRTHSGSDGSHSGQEGANTTSGPIRTPOIPMLIIILASLALIL 413

Qy 235 -----PSGSYTTTHASBSAGHSSGSEHGANVTSGPMKPGQIPMLIIILASLALIL 286

Db 235 -----PSGSYTTTHASBSAGHSSGSEHGANVTSGPMKPGQIPMLIIILASLALIL 286

Qy 414 ILAVCIANVSRRCGQKKVYINSGNVAVEDRKPSGLNGEASKSQEMVHLVNTKSSSTPD 473

Db 414 ILAVCIANVSRRCGQKKVYINSGNVAVEDRKPSGLNGEASKSQEMVHLVNTKSSSTPD 473

Qy 287 ILAVCIANVSRRCGQKKVYINSGNVAVEDRKPSGLNGEASKSQEMVHLVNTKSSSTPD 346

Db 287 ILAVCIANVSRRCGQKKVYINSGNVAVEDRKPSGLNGEASKSQEMVHLVNTKSSSTPD 346

Qy 474 QPMTADETRLQNVDMKIGV 493

Db 474 QPMTADETRLQNVDMKIGV 493

Qy 347 QPMTADETRLQNVDMKIGV 366

Db 347 QPMTADETRLQNVDMKIGV 366

## RESULT 11

A35616  
T-cell surface glycoprotein CD44 - hamster  
C:Species: Crictetinae gen. sp. (hamster)  
C/Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 21-Jul-2000  
C/Accession: A35616  
R:Aruffo, A.; Stamenkovic, I.; Melnick, M.; Underhill, C.B.; Seed, B.  
Cell 61, 1303-1313, 1990  
A:Title: CD44 is the principal cell surface receptor for hyaluronate.  
A:Reference number: A35616; MUID:90304889; PMID:1694723  
A:Accession: A35616  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Superfamily: human cell adhesion protein CD44  
C:Keywords: cell adhesion; glycoprotein; transmembrane protein

## Query Match

Best Local Similarity 54.7%; Score 1427.5; DB 2; Length 362;  
Matches 297; Conservative 23; Mismatches 37; Indels 141; Gaps 7;

Qy 1 MDKFWMAHMGCLVPLSLA--QIDNITCRFAGVFHVEKNGRYSISRTAADLCKAFNS 58

Db 1 MDKFWMAHMGCLVPLSLA--QIDNITCRFAGVFHVEKNGRYSISRTAADLCKAFNS 60

Qy 59 TLPTMAQMEKALSIGFETCRGFIEGHVVIRIHPNSICANNNGVYIITNTSGYDTYC 118

Db 59 TLPTMAQMEKALSIGFETCRGFIEGHVVIRIHPNSICANNNGVYIITNTSGYDTYC 120

Qy 119 FNASAPBEDCTSVTDLPLNAFDGPTITIVNRDGRVYVKGERTNEDIPSPNPTDDV 178

Db 119 FNASAPBEDCTSVTDLPLNAFDGPTITIVNRDGRVYVKGERTNEDIPSPNPTDDV 180

Qy 179 SSGSSERSSTISGCI-FYFT-FSTVHPIDEDSPWITDSTRIPATNMDSHSTTLQPTAN 237

Db 179 SSGSSERSSTISGCI-FYFT-FSTVHPIDEDSPWITDSTRIPATNMDSHSTTLQPTAN 237

Qy 181 SSGSSERSSTISGCI-FYFT-FSTVHPIDEDSPWITDSTRIPATNMDSHSTTLQPTAN 218

Db 181 SSGSSERSSTISGCI-FYFT-FSTVHPIDEDSPWITDSTRIPATNMDSHSTTLQPTAN 218

Qy 228 PNTGLVEDLDRGTPLSMITQOONSQSFSTHEGLEEDKHPTSTILSNRNDVYGGRD 297

Db 228 PNTGLVEDLDRGTPLSMITQOONSQSFSTHEGLEEDKHPTSTILSNRNDVYGGRD 297

Qy 219 -----MATR----- 222

Db 219 -----MATR----- 222

Qy 228 PNHSEGSTLLLEGYTSHPHTKESRTPIPVTSAKTGSFGVAVTVGDSNVNRSLSGD 357

Db 228 PNHSEGSTLLLEGYTSHPHTKESRTPIPVTSAKTGSFGVAVTVGDSNVNRSLSGD 357

Qy 223 ----- 224

Db 223 ----- 224

Qy 358 D-TFHPSGSHT-TGSHRSDGSHSGQEGANTTSGPIRTPOIPMLIIILASLALIL 415

Db 358 D-TFHPSGSHT-TGSHRSDGSHSGQEGANTTSGPIRTPOIPMLIIILASLALIL 415

Qy 225 DSSMDPRGNSLTVIDGSKLTGHSNGDGSANTTSRPRKQIPEMLIIILASLALIL 284

Db 225 DSSMDPRGNSLTVIDGSKLTGHSNGDGSANTTSRPRKQIPEMLIIILASLALIL 284

Qy 416 AVCIANVSRRCGQKKVYINSGNVAVEDRKPSGLNGEASKSQEMVHLVNTKSSSTPD 475

Db 416 AVCIANVSRRCGQKKVYINSGNVAVEDRKPSGLNGEASKSQEMVHLVNTKSSSTPD 475

Qy 285 AVCIANVSRRCGQKKVYINSGNVAVEDRKPSGLNGEASKSQEMVHLVNTKSSSTPD 344

Db 285 AVCIANVSRRCGQKKVYINSGNVAVEDRKPSGLNGEASKSQEMVHLVNTKSSSTPD 344

Qy 476 QPMTADETRLQNVDMKIGV 493

Db 476 QPMTADETRLQNVDMKIGV 493

Qy 345 QPMTADETRLQNVDMKIGV 362

Db 345 QPMTADETRLQNVDMKIGV 362

## RESULT 12

B38745  
cell adhesion molecule CD44 precursor, long form (meta-1) - rat  
C:Species: Rattus norvegicus (Norway rat)  
C/Date: 24-Jan-1992 #sequence\_revision 24-Jan-1992 #text\_change 05-Nov-1999  
C/Accession: B38745; A58745  
R:Guenther, U.; Hofmann, M.; Rudy, W.; Reber, S.; Zoeller, M.; Haubmann, I.; Matzke  
Cell 65, 13-24, 1991  
A:Title: A new variant of glycoprotein CD44 confers metastatic potential to rat car

A:Reference number: B38745; MUID:91191552; PMID:1707342  
A:Accession: B38745  
A:Status: preliminary  
A:Molecule type: mRNA



A:Residues: 1-503 <GUB>  
 A:Cross-references: GB:M61874; NID:g576534; PIDN:AAA53534.1; PID:g576535  
 A:Accession: A38745  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-223,386-503 <GU2>  
 A:Cross-references: GB:M61875  
 C:Keywords: cell adhesion

Query Match 54.6%; Score 1425; DB 2; Length 503;  
 Best Local Similarity 58.7%; Pred. No. 3.6e-84;  
 Matches 315; Conservative 36; Mismatches 108; Indels 78; Gaps 16;

```

Qy 1 MDKFWHAAWGLCLVPLSLA--QIDNITCRFAGVHVKNGRYSISRTAADLCKAF 57
Db 1 MDKFWHAAWGLCLVPLSLAQQIDNITCRFAGVHVKNGRYSISRTAADLCKAF 60
Qy 58 STLPYMAQMEKALSIGETCRYGFEIGHVVIPIRHNSICANNNGVYI-LTSNTSOYDT 116
Db 61 TLPTMAQMEALARKGETCRYGFEIGHVVIPIRHNAICANNNGVYIILASNTSHYDT 120
Qy 117 YCFNASAPPEEDCTSVTDLPNAPFGPITITVNRDGRVYQKGYRNPEDVPSNPTD 176
Db 121 YCFNASAPLEEDCTSVTDLPNSFGPVTITVNRDGRVYQKGRITHQEDLASNIID 180
Qy 177 DVSSGSSERSSTGGYIFYT-FSTVHPIDED-----SPWITDST---D 217
Db 181 DVSSGSTIEK-STEGYIILHTDLPSTQPTGDRDAFPIGSTLATIATTPVSAHTQNG 239
Qy 218 RIRATNMDSSHSTLQSTANPNGLVEDLR--TGPIASMTQOSQSPSTHSEGLEEK 275
Db 240 RLCQNPFIHSPVILQTTTR---MTDIDRNSTSAGENWTEPOPPFNHNEYQDEEST 294
Qy 276 DHPITSTLTSNRNDVYG-----GRDPNHSSESTLLLEGYT--SHYPTKE 320
Db 295 PHATSTWADPNSTTEBAATQKEKFNEMWQKNPPTSPDS-HVTEGTTASANNHPQ 353
Qy 321 SRITPIYTSKATGSGVTAATVGDNSNNRSLSGDDDT---HSGSGHTHGSBSOG 376
Db 354 RMT-----TQSGEDVSMT-----DFPDPIHWPQGHQF--ESKG 386
Qy 377 HSHSGQEGANTTSGPIRTPOIPFWILLIILASLALAILAVCIANSRRRCQKKLVIN 436
Db 387 HSGNGDSGVTTTSGPARPQIPFWILLIILASLALAILAVCIANSRRRCQKKLVIN 446
Qy 437 SGNGAVDRKPSGANGASQSQEWHLVKNESSETPDQFNTADETRNLQNVDMKIGV 493
Db 447 SGNGTVDRKPSSEINGASQSQEWHLVKNKEPETPDQFNTADETRNLQNVDMKIGV 503

```

RESULT 13  
 A37009  
 CD44 homolog membrane glycoprotein precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 31-Jan-1992 #sequence\_revision 31-Jan-1992 #text\_change 21-Jul-2000  
 C:Accession: A37009  
 R/Zhou, D.F.H.; Ding, J.F.; Pickler, L.J.; Bargatzke, R.F.; Butcher, E.C.; Goeddel, D.V.  
 J. Immunol. 143, 3390-3395, 1989  
 A:Title: Molecular cloning and expression of Bgp-1, The mouse homolog of the human H-CAM  
 A:Reference number: A37009; MUID:90038499; PMID:2681416  
 A:Accession: A37009  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-363 <ZHO>  
 A:Cross-references: GB:M30655; NID:g200332; PIDN:AAA3922.1; PID:g200333  
 C:Superfamily: human cell adhesion protein CD44  
 C:Keywords: cell adhesion; glycoprotein; membrane protein

Query Match 52.5%; Score 1368; DB 2; Length 363;  
 Best Local Similarity 58.1%; Pred. No. 1.1e-80;  
 Matches 289; Conservative 28; Mismatches 42; Indels 138; Gaps 9;

Qy 1 MDKFWHAAWGLCLVPLSLA--QIDNITCRFAGVHVKNGRYSISRTAADLCKAF 58

```

Db 1 MDKFWHAAWGLCLVPLSLAQQIDNITCRFAGVHVKNGRYSISRTAADLCKAF 60
Qy 59 TLPTMAQMEKALSIGETCRYGFEIGHVVIPIRHNSICANNNGVYI-LTSNTSOYDT 117
Db 61 TLPTMDQKALSKGFETCRYGFEIGHVVIPIRHNAICANNNGVYIILASNTSHYDT 120
Qy 118 CFNASAPPEEDCTSVTDLPNAPFGPITITVNRDGRVYQKGYRNPEDVPSNPTD 177
Db 121 CFNASAPPEEDCTSVTDLPNSFGPVTITVNRDGRVYQKGRITHQEDLASNIID 180
Qy 178 VSSGSSERSSTGGYIFYTFSTVHPIDEDSPWITDSTRIPATNMDSSHSTLQSTAN 237
Db 181 VSSGSTIEK-STEGYIILHTY-----LPTS-----QFPG- 208
Qy 238 PNTGLVEDLDRGTPLSMTQOSQSPSTHSEGLEEDKHDTTSTLTSNRNDVGGARD 297
Db 209 -----DQDSFFIRSTLARDSDS----- 227
Qy 298 PNHSEGSTLLLEGYTSYHTKESRTPIYTSKATGSGVTAATVGDNSNNRSLSGDQ 357
Db 228 -----SKDSR----- 232
Qy 358 DTFHPSGSGSH-TGSSSDGSHSGQEGANTTSGPIRTPOIPFWILLIILASLALILA 416
Db 233 -----GSSRTVTGSESLAGHSSANQDSGVTTTSGPMRFPQIPFWILLIILASLALILA 286
Qy 417 VCIANSRRRCQKKLVINGANGAVEDRKPSGANGASQSQEWHLVKNESSETPDQFM 476
Db 287 VCIANSRRRCQKKLVINGANGVEDRKPSSEINGASQSQEWHLVKNKEPETPDQFM 346
Qy 477 TADETRNLQNVDMKIGV 493
Db 347 TADETRNLQNVDMKIGV 363

```

RESULT 14  
 A34424  
 Cd44 membrane glycoprotein precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 15-Jun-1990 #sequence\_revision 15-Jun-1990 #text\_change 01-Dec-2000  
 C:Accession: A34424; A34907  
 R/Nottebohm, C.; Rees, G.; St. John, T.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 8521-8525, 1989  
 A:Title: Isolation of mouse Cd44 cDNA: structural features are distinct from the pri  
 A:Reference number: A34424; MUID:90046829; PMID:2682651  
 A:Accession: A34424  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-365 <NOT>  
 A:Cross-references: GB:M27130; NID:G192530; PIDN:AAA37407.1; PID:g309161  
 R/Wolfe, E.J.; Gause, W.C.; Pelfrey, C.M.; Holland, S.M.; Steinberg, A.D.; August,  
 U. Biol. Chem. 265, 341-347, 1990  
 A:Title: The cDNA sequence of mouse Bgp-1 and homology to human CD44 cell surface ar  
 A:Reference number: A34907; MUID:90044420; PMID:2403559  
 A:Accession: A34907  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 8-195, 'G', 197-365 <MOU>  
 A:Cross-references: GB:J05163; NID:g200334; PIDN:AAA3923.1; PID:g200335  
 C:Superfamily: human cell adhesion protein CD44  
 C:Keywords: cell adhesion; glycoprotein; membrane protein

Query Match 52.2%; Score 1361; DB 2; Length 365;  
 Best Local Similarity 57.7%; Pred. No. 3e-80;  
 Matches 288; Conservative 28; Mismatches 43; Indels 140; Gaps 9;

```

Qy 1 MDKFWHAAWGLCLVPLSLA--QIDNITCRFAGVHVKNGRYSISRTAADLCKAF 56
Db 1 MDKFWHAAWGLCLVPLSLAHPQIDNITCRFAGVHVKNGRYSISRTAADLCKAF 60
Qy 57 NSTLPYMAQMEKALSIGETCRYGFEIGHVVIPIRHNSICANNNGVYI-LTSNTSOYD 115

```

```
Db 61 NSTLPTMDQKTLAKSGFETCRVYFIEGNVVIPIHNAICANHTGVILVTSNTSHYD 120
QY 116 TYCFNAGAPPEEDCTSVTDLPNAPDGPITITIVNRDGTTRYVQGEYRTNEDIYPSNPTD 175
Db 121 TYCFNAGAPPEEDCTSVTDLPNSPDGPVITITIVNRDGTTRYVQGEYRTNEDIYPSNPTD 180
QY 176 DDVSSGSSSRSSTSGVIFFTFSTVHIPEDSPWITDSTDRIPATNMDSHSTLQPT 235
Db 181 DDVSSGSTIEK-STPESTYIHTY-----LPT-----QPT 209
QY 236 ANPNTGLVEDLDRTGPLSMTTQGSNSQSFSTSHGLEEDKHPTTSTLSSNNDVTGAR 295
Db 210 G-----DQDSFFIRSTLIATRDSDS-----229
QY 296 RDPNHSSESTLLBGTSHYPTKESRFFIPVTSAKTSGFVTAATVGDNSNVNRSLSG 355
Db 230 -----SKDSR-----234
QY 356 DQDTFHPSGGSHT-THGSESDGSHSGQEGGANTTSGPIRTPQIPEWLIILASLALALI 414
Db 235 -----GSRRTVTHGSELHGHSSANDSGVTTTSGPMRRPQIPEWLIILASLALALI 286
QY 415 LAVCIANVRRRCQKKLVNNGAVEDRKPSGLNGASKSGQEWHLVNXKESSETPDQ 474
Db 287 LAVCIANVRRRCQKKLVNNGAVEDRKPSGLNGASKSGQEWHLVNXKESSETPDQ 346
QY 475 FMTADETNLQNVDMKIGV 493
Db 347 CMTADETNLQNVDMKIGV 365
```

## RESULT 15

```
S24222
CD44 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
C:Accession: S24222
R:Hotmann, M.; Rudy, W.; Zoeller, M.; Toeig, C.; Ponta, H.; Herrlich, P.; Guenther, U.
Cancer Res. 51, 5292-5297, 1991
A:Title: CD44 splice variants confer metastatic behavior in rats: homologous sequences
A:Reference number: S24222; MUID:92005448; PMID:1717145
A:Accession: S24222
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-337 <HOP>
A:Cross-references: EMBL:X62739; NID:937651; PIDN:CAA44602.1; PID:937652
```

Query Match 26.8%; Score 699; DB 2; Length 337;

Best Local Similarity 69.7%; Pred. No. 9.6e-38;

Matches 147; Conservative 10; Mismatches 18; Indels 36; Gaps 5;

```
QY 161 YRTNPEDIYPSNPTDDVSSGSSSRSSTSGVIFFTFSTVHIP-----PDEDSPWITD 214
Db 146 YRQTPRE-----DSHSTGTGAASGHTS-----HMQGRITPSPEDSGW-TD 186
QY 215 STDRIP-----ATNMDSHSTLQPTANPTGLVEDLDRTGPLSMTTQGSNSQS 263
Db 187 FFNPISHPMGRGHQGRMDWDSSHTLQPTANPTGLVEDLDRTGPLSMTTQGSNSQS 246
QY 264 FSTSHGLEEDKHPTTSTLSSNNDVTGARDPNHSSESTLLBGTSHYPTKESRT 323
Db 247 FSTSHGLEEDKHPTTSTLSSNNDVTGARDPNHSSESTLLBGTSHYPTKESRT 306
QY 324 FIPVTSAKTSGFVTAATVGDNSNVNRSLS 354
Db 307 FIPVTSAKTSGFVTAATVGDNSNVNRSLS 337
```

Search completed: March 8, 2004, 06:12:10  
Job time : 24 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2004, 06:08:16 / Search time 17 Seconds  
(without alignments)

1510.034 Million cell updates/sec

Title: PCT-US01-51014-1

Perfect score: 2608  
Sequence: 1 MDKFWHAAAGLCLVPLSLA.....QPMADETRNLCNVDKIGV 493

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2468.5	94.7	742	1 CD44_HUMAN	P16070 h cd44 anti
2	1698.5	65.1	362	1 CD44_PAPHA	P14745 papio hamad
3	1662.5	63.7	778	1 CD44_MOUSE	P15379 mus musculu
4	1631	62.5	431	1 CD44_MESAU	Q60522 m cd44 anti
5	1547	59.3	359	1 CD44_HORSE	Q05078 equus cabal
6	1470	56.4	351	1 CD44_CANPA	Q28284 canis famli
7	1469.5	56.3	366	1 CD44_BOVIN	Q29423 bos tauris
8	1425	54.6	503	1 CD44_RAT	P28051 rattus norv
9	1398.5	53.6	362	1 CD44_CRIGR	P20944 cricetulus
10	177	6.8	389	1 SER1_BOMMO	P07856 bombyx mori
11	171	6.6	276	1 TS66_RABIT	P98065 oryctolagus
12	168	6.4	2124	1 PGCA_RAT	P07897 rattus norv
13	163	6.2	275	1 TS66_MOUSE	Q08859 mus musculu
14	162.5	6.2	277	1 TS66_HUMAN	Q02496 mus musculu
15	158	6.1	630	1 MUC1_MOUSE	Q61882 mus musculu
16	155.5	6.0	2132	1 PGCV_MOUSE	Q62059 mus musculu
17	154	5.9	3358	1 PGCV_MOUSE	Q62059 mus musculu
18	152.5	5.8	3381	1 PGCV_BOVIN	P81882 bos tauris
19	151.5	5.8	2333	1 PGCA_CANPA	Q28343 canis famli
20	151	5.8	2109	1 PGCA_CHICK	P07898 gallus galli
21	150.5	5.8	636	1 YN6E_YEAST	P53882 saccharomyc
22	150	5.8	862	1 PGCV_MACNE	Q28858 macaca neme
23	148	5.7	1357	1 AMRH_YEAST	P08640 saccharomyc
24	147	5.6	3178	1 YS89_CAEEL	Q09624 caenorhabdi
25	146	5.6	2415	1 PGCA_HUMAN	P16112 homo sapien
26	145	5.6	883	1 PGCB_RAT	P55068 rattus norv
27	143.5	5.5	1306	1 MSB2_YEAST	P32334 saccharomyc
28	141.5	5.4	937	1 HYRI_CANAL	P46591 candida alb
29	141	5.4	797	1 VGLX_HSVB	Q28868 equine hepr
30	139	5.3	537	1 PGCA_PIG	Q29011 sus scrofa
31	139	5.3	681	1 VGP_YABVP	P35254 martus vir
32	139	5.3	1161	1 DAN4_YEAST	P47179 saccharomyc
33	139	5.3	1260	1 AL51_CANAL	P46590 candida alb

34	139	5.3	3396	1 PGCV_HUMAN	P13611 homo sapien
35	137	5.3	1169	1 YK82_YEAST	P36170 saccharomyc
36	136.5	5.2	1253	1 DSPF_HUMAN	Q02104 homo sapien
37	135	5.2	907	1 VGP3_EBV	P03200 Epstein-Barr
38	134	5.1	3562	1 PGCV_CHICK	Q90953 gallus galli
39	133.5	5.1	556	1 MSC3_YEAST	Q12215 saccharomyc
40	133.5	5.1	2316	1 PTP2_RAT	Q62656 rattus norv
41	133.5	5.1	670	1 YFEG_SCHPO	Q13854 schistosach
42	133	5.1	1257	1 PGCV_RAT	P55067 rattus norv
43	133	5.1	5703	1 MUS8_HUMAN	Q9hc84 homo sapien
44	132.5	5.1	1403	1 YDF3_SCHPO	Q10475 schistosach
45	132	5.1	1075	1 FLOS_YEAST	P38894 saccharomyc

## ALIGNMENTS

RESULT 1  
CD44\_HUMAN STANDARD; PRT; 742 AA.  
ID CD44\_HUMAN  
AC P16070; P22511; Q04858; Q13419; Q13957; Q13958; Q13959; Q13960;  
AC Q13961; Q13967; Q13968; Q13980; Q15861; Q16064; Q16065; Q16066;  
AC Q16208; Q16522; Q96024;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 15-MAR-2004 (Rel. 43, Last annotation update)  
DE CD44 antigen precursor (Phagocytic glycoprotein I) (PGP-I) (HUTCH-I)  
DE (Extracellular matrix receptor-II) (ECMR-II) (GP90 lymphocyte  
DE homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor)  
DE (Heparan sulfate proteoglycan) (Epican) (CDw44).  
GN CD44 OR LHR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=Lymphoblast;  
RX MEDLINE=93101687; PubMed=1465456;  
RA Screation G.R., Bell M.V., Jackson D.G., Cornelis F.B., Gerth U.,  
Bell J.I.;  
RT "Genomic structure of DNA encoding the lymphocyte homing receptor  
CD44 reveals at least 12 alternatively spliced exons.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:12160-12164(1992).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM EPIDERMAL).  
RC TISSUE=Keratinocytes;  
RA Kugelman L.C., Ganguly S., Haggerty J.G., Weissman S.M.,  
Milstone L.M.;  
RT "The core protein of epican, a heparan sulfate proteoglycan on  
keratinocytes, is an alternative form of CD44.";  
RL J. Invest. Dermatol. 99:381-385(1992).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM EPITHELIAL CD44E).  
RC MEDLINE=91122041; PubMed=1991450;  
RA Stamenkovic I., Arnoff A., Amiot M., Seed B.;  
RT "The hematopoietic and epithelial forms of CD44 are distinct  
polypeptides with different adhesion potentials for  
hyaluronate-bearing cells.";  
RL EMBO J. 10:343-348(1991).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORMS CD44R1 AND CD44R2).  
RC TISSUE=Myeloid leukemia cells;  
RX MEDLINE=91277598; PubMed=2056274;  
RA Dougherty G.J., Lansdorp P.M., Cooper D.L., Humphries R.K.;  
RT "Molecular cloning of CD44R1 and CD44R2, two novel isoforms of the  
human CD44 lymphocyte 'homing' receptor expressed by hemopoietic  
cells.";  
RL J. Exp. Med. 174:1-5(1991).  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORMS CD44SP, CD44R4 AND CD44R5).  
RC TISSUE=Breast carcinoma;  
RX MEDLINE=93356912; PubMed=8352881;

RA Tanabe K.K., Nishi T., Sava H.;  
RT "Novel variants of CD44 arising from alternative splicing: changes in  
RT the CD44 alternative splicing pattern of MCF-7 breast carcinoma cells  
RT treated with hyaluronidase.";  
RL Mol. Carcinog. 7:212-220(1993).  
[6]  
RC SEQUENCE FROM N.A. (ISOFORM RETICULOCYTE).  
RC TISSUE=Reticulocytes;  
RX MEDLINE=91337049; PubMed=1840487;  
RA Hahn H.J., Isola N., Cooper D.L.;  
RT "The multispecific cell adhesion molecule CD44 is represented in  
RT reticulocyte cDNA.";  
RL Biochem. Biophys. Res. Commun. 178:1127-1134(1991).  
[7]  
RC SEQUENCE FROM N.A.  
RX MEDLINE=89168434; PubMed=2466575;  
RA Stamenkovic I., Amiot M., Pesando J.M., Seed B.;  
RT "A lymphocyte molecule implicated in lymph node homing is a member of  
RT the cartilage link protein family.";  
RL Cell 56:1057-1062(1989).  
[8]  
RC SEQUENCE FROM N.A. (ISOFORMS SHORT-TAILED HEMATOPOIETIC AND CD44H).  
RX MEDLINE=89168435; PubMed=2466576;  
RA Goldstein L.A., Zhou D.F.H., Pickar L.J., Minty C.N., Bargatze R.F.,  
RA Ding J.F., Butcher E.C.;  
RT "A human lymphocyte homing receptor, the Hermes antigen, is related  
RT to cartilage proteoglycan core and link proteins.";  
RL Cell 56:1063-1072(1989).  
[9]  
RC SEQUENCE FROM N.A. (ISOFORM WITHOUT EXON 6).  
RC TISSUE=Pancreas;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschuler S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toehlyuki S., Abramson R.D., Mullaly S.J.,  
RA Rana S.S., Loguélano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.V., Hulyk S.W.,  
RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Paley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
RA Schermer A., Schein J.E., Jones S.J.M., Marra W.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[10]  
RC SEQUENCE OF 184-222 AND 473-625 FROM N.A. (ISOFORM KERATINOCYTE).  
RC TISSUE=Fore skin;  
RX MEDLINE=91177958; PubMed=2007624;  
RA Brown T.A., Bouchard T., St John T., Wayner E., Carter W.G.;  
RT "Human keratinocytes express a new CD44 core protein (CD44S) as a  
RT heparan-sulfate intrinsic membrane proteoglycan with additional  
RT exons.";  
RL J. Cell Biol. 113:207-221(1991).  
[11]  
RC SEQUENCE FROM N.A.  
RC TISSUE=Articular cartilage;  
RX Boshch P.P., Stevens J.W., Buckwalter J.A., Midura R.J.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
[12]  
RC SEQUENCE OF 267-603 FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=92005448; PubMed=1717145;  
RA Hofmann M., Rudy W., Zoeller M., Toelg C., Ponta H., Herrlich P.,  
RA Guenther U.;  
RT "CD44 splice variants confer metastatic behavior in rats: homologous

RT sequences are expressed in human tumor cell lines.";  
RL Cancer Res. 51:5292-5297(1991).  
[13]  
RC SEQUENCE OF 223-265 FROM N.A.  
RX MEDLINE=94198700; PubMed=8146709;  
RA Matsumura Y., Hanbury D., Smith J., Tatin D.;  
RT "Non-invasive detection of malignancy by identification of unusual  
RT CD44 gene activity in exfoliated cancer cells.";  
RL BMC 308:619-624(1994).  
[14]  
RC SEQUENCE OF 1-22 FROM N.A.  
RC TISSUE=Lymphoblast;  
RX MEDLINE=92017823; PubMed=1922057;  
RA Shivelman E., Bishop J.M.;  
RT "Expression of CD44 is repressed in neuroblastoma cells.";  
RL Mol. Cell. Biol. 11:5446-5453(1991).  
[15]  
RC REVIEW ON FUNCTION AND POST-TRANSLATIONAL MODIFICATIONS.  
RX MEDLINE=22399881; PubMed=12511867;  
RA Ponta H., Sherman L., Herrlich P.A.;  
RT "CD44: from adhesion molecules to signalling regulators.";  
RL Nat. Rev. Mol. Cell Biol. 4:33-45(2003).  
[16]  
RC PHOSPHORYLATION OF SER-706.  
RX MEDLINE=98248445; PubMed=9580567;  
RA Peck D., Isaacke C.M.;  
RT "Hyaluronan-dependent cell migration can be blocked by a CD44  
RT cytoplasmic domain peptide containing a phosphoserine at position  
RT 325.";  
RL J. Cell Sci. 111:1595-1601(1998).  
[17]  
RC PHOSPHORYLATION OF SER-672.  
RX MEDLINE=22038351; PubMed=12032545;  
RA Legg J.W., Lewis C.A., Parsons M., Ng T., Isaacke C.M.;  
RT "A novel PKC-regulated mechanism controls CD44 ezrin association and  
RT directional cell motility.";  
RL Nat. Cell Biol. 4:399-407(2002).  
[18]  
RC GLYCOSYLATION AND PROCESSING.  
RX MEDLINE=22766015; PubMed=12883358;  
RA Bartolazzi A.;  
RT "CD44 adhesion function spontaneous and PMA-inducible CD44 cleavage  
RT are regulated at post-translational level in cells of melanocytic  
RT lineage.";  
RL Melanoma Res. 13:325-337(2003).  
[19]  
RC VARIANT BLOOD GROUP INDIAN PRO-46.  
RX MEDLINE=96215152; PubMed=8636151;  
RA Telen M.J., Udani M., Washington M.K., Levesque M.C., Lloyd E.,  
RA Rao N.;  
RT "A blood group-related polymorphism of CD44 abolishes a hyaluronan-  
RT binding consensus sequence without preventing hyaluronan binding.";  
RL J. Biol. Chem. 271:7147-7153(1996).  
[20]  
RC -1- FUNCTION: Receptor for hyalurononic acid (HA). Mediates cell-cell  
and cell-matrix interactions through its affinity for HA, and  
possibly also through its affinity for other ligands such as  
osteopontin, collagen, and matrix metalloproteinases (MMPs).  
Adhesion with HA plays an important role in cell migration, tumor  
growth and progression. Also involved in lymphocyte activation,  
recirculation and homing, and in hematopoiesis. Altered expression  
or dysfunction causes numerous pathogenic phenotypes. Great  
protein heterogeneity due to numerous alternative splicing and  
post-translational modification events.  
[21]  
RC -1- SUBUNIT: Interacts with HA, as well as other glycosaminoglycans,  
collagen, laminin, and fibronectin via its N-terminal segment.  
Interacts with ANK, the ERM proteins (VIL2, RDX and  
MSN), and NF2 via its C-terminal segment.  
[22]  
RC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
[23]  
CC -1- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Comment=Additional isoforms seem to exist. Additional isoforms  
are produced by alternative splicing of 10 out of 19 exons  
within the extracellular domain. Additional diversity is

Query Match 94.7%; Score 2468.5; DB 1; Length 742;  
 Best Local Similarity 66.3%; Pred. No. 2.5e-146;  
 Matches 492; Conservative 1; Mismatches 0; Indels 249; Gaps 1;

QY 1 MDKFWMAAMGCLVPLSLAQLDINTCRFAGVHYEKNGRYSISRTEADLCKAFNSTL 60  
 DB 1 MDKFWMAAMGCLVPLSLAQLDINTCRFAGVHYEKNGRYSISRTEADLCKAFNSTL 60  
 QY 61 PTMAQWEKALSIGFETCRYGFLGHVAVIRIHNISCAANNNGVYLINSQYDYTCFN 120  
 DB 61 PTMAQWEKALSIGFETCRYGFLGHVAVIRIHNISCAANNNGVYLINSQYDYTCFN 120  
 QY 121 ASAPREDCTSVLDLNAFDGPITITVNRDGRYVOKGEYRNTNEDIPSNPTDDVSS 180  
 DB 121 ASAPREDCTSVLDLNAFDGPITITVNRDGRYVOKGEYRNTNEDIPSNPTDDVSS 180  
 QY 181 GSSERSSTGGYITFTFTSTVHPIDPESDPWITDSTRIPAT----- 222  
 DB 181 GSSERSSTGGYITFTFTSTVHPIDPESDPWITDSTRIPAT----- 222  
 QY 223 ----- 222  
 DB 223 ----- 222  
 QY 241 TMDWFWLPLPSKXHLHTTOMAGTSNTISAGWPEWENEDERDLHSFGSGIDDD 300  
 DB 241 TMDWFWLPLPSKXHLHTTOMAGTSNTISAGWPEWENEDERDLHSFGSGIDDD 300  
 QY 223 ----- 222  
 DB 301 EDFISSTITTPPAFHTKQNDPMTQWNSHNEVLQTTMTMDVDENGTAYEGWN 360  
 QY 223 ----- 222  
 DB 361 PEAPHPLIHHEHEEETPHSTTIOATPSSTTEATQKQFQKRWHEGRQTPRED 420  
 QY 223 ----- 222  
 DB 421 HSTTGAASARTSHPMQGRTPSPEDSWITDFNFISHPMGRHQAGRMMDSHSHT 480  
 QY 232 LQFTANPNGLVLDLRTGPLMTQOONSQSFSTHGLKEDKHPTSTLTSNRNDV 291  
 DB 481 LQFTANPNGLVLDLRTGPLMTQOONSQSFSTHGLKEDKHPTSTLTSNRNDV 291  
 QY 292 TGGRRDPNHSSTLLLEGYTSHYPTKSRFIPVTSKTSFGVAVAYVDSNPNVR 351  
 DB 541 TGGRRDPNHSSTLLLEGYTSHYPTKSRFIPVTSKTSFGVAVAYVDSNPNVR 351  
 QY 352 SLSGDDPTFRPSGSHHTGSESDGSHSQSDEGANTTGPRTPIPEWLLIILASLAL 411  
 DB 601 SLSGDDPTFRPSGSHHTGSESDGSHSQSDEGANTTGPRTPIPEWLLIILASLAL 411  
 QY 412 ALLIACIAVNSRRRGQKKQLVINSNGAVEDRKPSGLNGEASKSQEWVHLNKESSST 471  
 DB 661 ALLIACIAVNSRRRGQKKQLVINSNGAVEDRKPSGLNGEASKSQEWVHLNKESSST 471  
 QY 472 PDQFTADERTANQNDMKIGV 493  
 DB 721 PDQFTADERTANQNDMKIGV 742

RESULT 2  
 CD44\_PAPHA STANDARD; PRT; 362 AA.  
 ID P14745;  
 AC 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE CD44 antigen precursor (Phagocytic glycoprotein I) (PC9-1) (HUTCH-1)  
 DE (Extracellular matrix receptor-III) (ECMR-III) (GP90 lymphocyte  
 DE homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor).  
 GN CD44.  
 OS Papio hamadryas (Hamadryas baboon).  
 OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Papio.

NCBI\_TaxID=9557;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 66-74.  
 RX MEDLINE=69282830; PubMed=2471974;  
 RA Idzerda R.L., Carter W.G., Nottenburg C., Wayner E.A.,  
 RA Gallatin W.M., St John T.J.  
 RT "Isolation and DNA sequence of a cDNA clone encoding a lymphocyte  
 RT adhesion receptor for high endothelium."  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4659-4663 (1989).  
 CC -1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to  
 CC mucosal high endothelial venule and to types I and VI collagen.  
 CC Probably involved in matrix adhesion, lymphocyte activation and  
 CC lymph node homing.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- PTM: Extensively modified including N- and O-linked glycosylation,  
 CC addition of the glycosaminoglycan chondroitin sulfate, of sulfate,  
 CC of phosphate to cytoplasmic domain serine residues.  
 CC -1- SIMILARITY: Contains 1 link domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M22452; AAA35385.1; -.  
 DR HSSP: P98066; ITSG.  
 DR InterPro: IPR001231; CD44 antigen.  
 DR InterPro: IPR000538; Link.  
 DR Pfam: PF00193; Xlink; 1.  
 DR PRINTS: PR00658; CD44.  
 DR PRINTS: PR01265; LINKMODULE.  
 DR PRODOM: PD000918; Link; 1.  
 DR SMART: SM00445; Link; 1.  
 DR PROSITE: PS01241; Link; 1.  
 KW Cell adhesion; Transmembrane; Glycoprotein; Phosphorylation; Receptor;  
 KW Proteoglycan; Signal; Alternative splicing;  
 KM Pyroglutamate carboxylic acid.  
 FT SIGNAL 1 20  
 FT CHAIN 21 362  
 FT DOMAIN 21 269  
 FT TRANSMEM 270 290  
 FT DOMAIN 291 362  
 FT DOMAIN 48 119  
 FT DOMAIN 150 158  
 FT DOMAIN 225 269  
 FT MOD\_RES 21 21  
 FT DISULFID 53 118  
 FT DISULFID 77 97  
 FT MOD\_RES 292 292  
 FT MOD\_RES 326 326  
 FT CARBOHYD 25 25  
 FT CARBOHYD 57 57  
 FT CARBOHYD 100 100  
 FT CARBOHYD 110 110  
 FT CARBOHYD 120 120  
 FT CARBOHYD 256 256  
 FT CONFLICT 67 67  
 SQ SEQUENCE 362 AA; 39378 MW; 578BCE/C3D52EFF CRC64;

Query Match 65.1%; Score 1698.5; DB 1; Length 362;  
 Best Local Similarity 68.2%; Pred. No. 7.7e-99;  
 Matches 337; Conservative 7; Mismatches 17; Indels 133; Gaps 2;

QY 1 MDKFWMAAMGCLVPLSLAQLDINTCRFAGVHYEKNGRYSISRTEADLCKAFNSTL 60  
 DB 1 MDKFWMAAMGCLVPLSLAQLDINTCRFAGVHYEKNGRYSISRTEADLCKAFNSTL 60  
 QY 61 PTMAQWEKALSIGFETCRYGFLGHVAVIRIHNISCAANNNGVYLINSQYDYTCFN 120

Db 61 PTMAQWELKLSIGETRCRGVFIHGAVIPIHFNISCAANTGVILTSNTSGYDTCN 120  
 Qy 121 ASAPPEDCISYVDLPNAPFGPITITVNDGRIYQKGRYRNRPDIYPSNTPDDVS 180  
 Db 121 ASAPGBCDTSYVDLPNAPFGPITITVNDGRIYQKGRYRNRPDIYPSNTPDDVS 180  
 Qy 181 GSSSERSTSGYIFYT-FSTVAPIDPDSFWITDSTRIPATNMDSSHSTLQPTANEN 239  
 Db 181 GSSSERSTLGGYIFYFVNFHFSPIPEDGPWITDSTRIPATR----- 224  
 Qy 240 TGLVEDIDRTPGPIISMTTQGSNSGFSHSGLEDHDPITSTLTSSNNDVYGRDNN 299  
 Db 225 ----- 224  
 Qy 300 HSEGSTLLLEGYSHPHYTESRTFIPVTSAKTGSFGVAVTVGDSNMYNRLSGDQDT 359  
 Db 225 -----DQGA 228  
 Qy 360 FHPSGGSHTHGSSDGHSGSQEGANTTSGPRTQPIPEWIIILASLALALILANCI 419  
 Db 229 FDPGSGHHTHGSASHSGSREGANTTSGPLTPQIPEWIIILASLALALILAVCI 288  
 Qy 420 AVNSRRRCGQKKKIVINSNGAVEDRKPSSGLNGBASQGMVHLVKNSESETPDQMTAD 479  
 Db 289 AVNSRRRCGQKKKIVINSNGAVEDRKPSSGLNGBASQGMVHLVKNSESETPDQMTAD 348  
 Qy 480 ETRNLQNVDMKIGV 493  
 Db 349 ETRNLQNVDMKIGV 362  
 RESULT 3  
 CD44 MOUSE STANDARD; PRT; 778 AA.  
 ID CD44 MOUSE  
 AC P15379; Q05732; Q61395; Q62060; Q62061; Q62063; Q62408;  
 AC Q62409; Q05732; Q61395; Q62060; Q62061; Q62063; Q62408;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE CD44 antigen precursor (phagocytic glycoprotein I) (Pgp-1) (HUTCH-I)  
 DE (extracellular matrix receptor-II) (ECCR-II) (GP90 lymphocyte  
 DE homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor)  
 DE (LX-24).  
 GN CD44.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A. (ISOFORMS 4; 6; 7 AND 12).  
 RC STRAIN=DBA/2; TISSUE=Lung;  
 RC MEDLINE=93107170; PubMed=1469058;  
 RA He O., Lesley J., Hyman R., Ishihara K., Kinsade P.W.;  
 RA "Molecular isoforms of murine CD44 and evidence that the membrane  
 RT proximal domain is not critical for hyaluronate recognition.";  
 RL J. Cell Biol. 119:1711-1719(1992).  
 RN [2]  
 RN SEQUENCE FROM N.A. (ISOFORM 13).  
 RC MEDLINE=90038499; PubMed=2681416;  
 RA Zhou D.F.H., Ding J.F., Picker L.J., Bargatze R.F., Butcher E.C.,  
 RA Goeddel D.V.;  
 RA "Molecular cloning and expression of pgp-1. The mouse homolog of the  
 RT human H-CAM (Hermes) lymphocyte homing receptor.";  
 RL J. Immunol. 143:3390-3395(1989).  
 RN [3]  
 RN SEQUENCE FROM N.A. (ISOFORM 13).  
 RC MEDLINE=90046829; PubMed=2682511;  
 RA Notemburg C., Rees G., St John T.;  
 RA "Isolation of mouse CD44 cDNA: structural features are distinct from  
 RT the primate CDNA.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:8521-8525(1989).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=20318634; PubMed=10859330;  
 RA Witting B.M., Johansson B., Zoeller M., Schwaerzler C., Guentert U.;  
 RT "Abrogation of experimental colitis correlates with increased  
 RT apoptosis in mice deficient for CD44 variant exon 7 (CD44v7).";  
 RL J. Exp. Med. 191:2053-2064(2000).  
 RN [5]  
 RN SEQUENCE FROM N.A. (ISOFORM 13).  
 RP MEDLINE=22388257; PubMed=12477932;  
 RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Stensberg R.D., Collins F.S., Wagner L., Sherman C.M., Schuller G.D.,  
 RA Alechul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [6]  
 RN SEQUENCE FROM N.A. (ISOFORM 13).  
 RP STRAIN=C57BL/6J; TISSUE=Embryo;  
 RC MEDLINE=22354683; PubMed=12466851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nixardo I., Ose N., Saito R., Suzuki H., Yamana K.I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schobach C., Gojochi T.,  
 RA Balazs I.L., Hill D.P., Bult C., Hume D.A., Quackenbush U.,  
 RA Schmitt L.M., Kanapin A., Matsumura H., Batilov S., Beisel K.W.,  
 RA Blake J.A., Brad D., Brusic V., Chotcha C., Corbani L.B., Cousins S.,  
 RA Dalia E., Dragani T.A., Fletcher C.F., Forrest A., Pazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Gough J.,  
 RA Grimmond S., Guerinich S., Hirokawa N., Jackson I.T., Jarvis E.D.,  
 RA Karai A., Kawaji H., Kawasawa Y., Kedziarski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lehner B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Nunata K., Okido T., Pavan W.J., Perte G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Warande Y., Wells C.,  
 RA Wilmig L.G., Wynshaw-Boris A., Yangisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Komori H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [7]  
 RN SEQUENCE OF 8-778 FROM N.A. (ISOFORM 13).  
 RP MEDLINE=90094420; PubMed=2403559;  
 RA Wolfe E.U., Gause W.C., Peitrey C.M., Holland S.M., Steinberg A.D.,  
 RA August J.T.;  
 RT "The cDNA sequence of mouse Pgp-1 and homology to human CD44 cell  
 RT surface antigen and proteoglycan core/link proteins.";  
 RL J. Biol. Chem. 265:341-347(1990).  
 RN [8]  
 RP SEQUENCE OF 224-637 FROM N.A. (ISOFORMS 1; 2; 3; 4; 5; 6; 7 AND 8).  
 RC STRAIN=GR;

CC MEDLINE=93219085; PubMed=8464707;  
RA Toelg C., Hofmann M., Herrlich P., Ponta H.,  
RT "Splicing choice from ten variant exons establishes CD44  
RT variability.";  
RL Nucleic Acids Res. 21:1225-1229(1993).  
RN [9]  
RP SEQUENCE OF 224-637 FROM N.A. (ISOFORM 9).  
RC STRAIN=BALB/c;  
RX MEDLINE=93286043; PubMed=8509359;  
RA Screaton G.R., Bell M.V., Bell J.I., Jackson D.G.,  
RT "The identification of a new alternative exon with highly restricted  
RT tissue expression in transcripts encoding the mouse Pgp-1 (CD44)  
RT homing receptor. Comparison of all 10 variable exons between mouse,  
RT human, and rat.";  
RL J. Biol. Chem. 268:12235-12238(1993).  
RN [10]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 10 AND 11).  
RC STRAIN=Swiss Webster;  
RX MEDLINE=96355396; PubMed=8702806;  
RA Yu Q., Toole B.P.,  
RT "A new alternatively spliced exon between v9 and v10 provides a  
RT molecular basis for synthesis of soluble CD44.";  
RL J. Biol. Chem. 271:20603-20607(1996).  
CC -1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to  
CC mucosal high endothelial venule and to types I and VI collagen.  
CC Probably involved in matrix adhesion, lymphocyte activation and  
CC lymph node homing.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=13;  
CC Name=1;  
CC IsoId=P15379-14; Sequence=Displayed;  
CC Name=2;  
CC IsoId=P15379-7; Sequence=VSP\_007329;  
CC Name=3;  
CC IsoId=P15379-8; Sequence=VSP\_007330;  
CC Name=4; Synonyms=M2;  
CC IsoId=P15379-4; Sequence=VSP\_007331;  
CC Name=5;  
CC IsoId=P15379-9; Sequence=VSP\_007332;  
CC Name=6; Synonyms=M3;  
CC IsoId=P15379-5; Sequence=VSP\_005326;  
CC Name=7; Synonyms=M4;  
CC IsoId=P15379-6; Sequence=VSP\_005327;  
CC Name=8;  
CC IsoId=P15379-10; Sequence=VSP\_007330; VSP\_007334;  
CC Name=9;  
CC IsoId=P15379-11; Sequence=VSP\_007332; VSP\_007335;  
CC Name=10;  
CC IsoId=P15379-12; Sequence=VSP\_007336; VSP\_007337;  
CC Name=11;  
CC IsoId=P15379-13; Sequence=VSP\_007338; VSP\_007339;  
CC Name=12; Synonyms=M1;  
CC IsoId=P15379-3; Sequence=VSP\_005328;  
CC Name=13; Synonyms=M0;  
CC IsoId=P15379-2; Sequence=VSP\_005329;  
CC -1- PTM: N-glycosylated (By similarity);  
CC -1- PTM: O-glycosylated; contains chondroitin sulfate glycans which  
CC can be more or less sulfated (By similarity).  
CC -1- PTM: Phosphorylated; activation of PKC results in the  
CC dephosphorylation of Ser-742 (constitutive phosphorylation site),  
CC and the phosphorylation of Ser-708 (By similarity).  
CC -1- POLYMORPHISM: Two allelic forms of this glycoprotein, PGP-1.1 and  
CC PGP-1.2, have been reported. The expressed product is PGP-1.1 (ly-  
CC 24.1).  
CC -1- SIMILARITY: Contains 1 link domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
CC DR EMBL; X66084; CAA46883.1; -  
CC DR EMBL; X66083; CAA46882.1; -  
CC DR EMBL; X66082; CAA46881.1; -  
CC DR EMBL; X66081; CAA46880.1; -  
CC DR EMBL; M30655; AAA39922.1; -  
CC DR EMBL; M27129; AAA37406.1; -  
CC DR EMBL; M27130; AAA37407.1; -  
CC DR EMBL; AJ251594; CAB61888.1; -  
CC DR EMBL; BC005676; AAH05676.1; -  
CC DR EMBL; AK045226; BAC32269.1; -  
CC DR EMBL; J05163; AAA39923.1; -  
CC DR EMBL; X69724; CAA49380.1; -  
CC DR EMBL; U13611; AAA37145.1; -  
Query Match 63.7%; Score 1662.5; DB 1; Length 778;  
Best Local Similarity 47.4%; Pred. No. 3; 6e-96;  
Matches 370; Conservative 39; Mismatches 82; Indels 289; Gaps 10;  
-----  
QY 1 MDKFMWHAAGLGVPLSLA--QIDNITCRFAGVFHVKNGRYSISRTAADLCKAFNS 58  
DB 1 MDKFMWHAAGLGVPLSLA--QIDNITCRFAGVFHVKNGRYSISRTAADLCKAFNS 60  
QY 59 TLPTMAQMEKALSIGFETCRVGFEGHVIPIRIHPSICAAANTGVYL-TSNTSQYDTY 117  
DB 61 TLPTMDQKALSKGFETCRVGFEGHVIPIRIHPSICAAANTGVYLVTNSHTYDTY 120  
QY 118 CFNASAPPEEDCTSVTLNPAFDGPITITVNPCTGYVKGERTNPEDIPENPD 177  
DB 121 CFNASAPPEEDCTSVTLNPAFDGPITITVNPCTGYVKGERTNPEDIPENPD 180  
QY 178 VSSGSSSESTSGCYLFYTP-----STVH----- 202  
DB 181 VSSGSTIEK-STPGYLITHTLPTPEQPTGDDSFIRSTLATIATVSKSHAQAKON 239  
QY 203 ----- 202  
DB 240 NWIWSFGNSQSTQTOGPTTSATVLTMTPTPKQEAQNFWSLPOPESSKSHHTT 299  
QY 203 ----- 202  
DB 300 TKMGTESNTNPTGMEPNEBNEDETDYPPSSGSGIDDDDEFISITATTPRVGARTEDN 359  
QY 203 ----- 202  
DB 360 QDWTQWKENHNPVLLQTTTRMADIDRISTSAHGENWTPPEPPFNNHEXODEEETPHA 419  
QY 203 -----PIPEDS----- 209  
DB 420 TSTTPNSTAAEAATQGETWFGONGQGNPPRPSDSHVITGCTTASANNHNSCRITQSQ 479  
QY 210 --PWITDSTDR-----DATNDSHSTTLQPTANPNTGLVEDLDRTPPLAMTT 256  
DB 480 EDVSW-TDFPDPISHPMQGHQTESKPTDSHSHTLPTAPFNHVLVEDLNRTPLPVTT 538  
QY 257 QQSNQSGSTSHGEGHEHDKHPTSTLTSNRNVGTGRDPNHSBSSTTLLEGYTHYP 316  
DB 539 POSHQNSTTHGPEEDENFTTSLIPSSKSKAKARKGSLPTDTTIVSEYTYQYF 598  
QY 317 HTKESRTPITPYSKATGSGYTAATVAG-DNSNVNRLSGDDQTFHPGSGSH--TTGSE 373  
DB 599 DTMENGLTFPTPAKTEVFGTEVTLTDSNVNVDGLPGDRDSKDSRGSSRTVTGSGE 658  
QY 374 SDGSHSGQEGAGNTTSGPRTPTPIPEMLIILASLALALILAVCIANVSRRCQKKL 433  
DB 659 LAGHSAAQDSGVTTTSGPMRPPPIPEMLIILASLALALILAVCIANVSRRCQKKL 718  
QY 434 VINGNGAVEDRKPGLNGEAKSKQEMVHLVKNKSSSTPDQFMATDETRNLQNVDMKIGV 493  
DB 719 VINGNGVVEVRKSEINGEAKSKQEMVHLVKNKSSSTPDQFMATDETRNLQNVDMKIGV 778





D	b		_MDKFFWRYAAAGLCLVPLSLAQIDINTLCRYAIVFHEVKKNRYSISRTAADLKCAFNSTL	60
Q	y		61 PTMAQNEKLSTIGFETCRVYFIIGHVVIPIIHNSI <sup>CAANTGVILT</sup> STSGDYCN	120
D	b		61 PTMAQMKALNIFETCRIGFIIGHVVIPIIHNSI <sup>CAANTGVILT</sup> STSGDYCN	120
Q	y		121 ASAPPEEDCTSVTDLPNAFDGPITITIVNRDGTTRYVGKEGRTNPEDIYPSEPTDDVSS	180
D	b		121 ASAPPEEDCTSVTDLPNAFDGPITITIVNRDGTTRYVGKEGRTNPEDINSTEADDDVSS	180
Q	y		181 GSSSERSTSGCY-IPT-PSYHAPIDEDSMTITSDTRIPATNNDSSTLLQPANP	238
D	b		181 GSSSER-STGGGSI <sup>SIFETHLPTR</sup> FDDQSFPVSDPEREPTT-	223
Q	y		239 NTGLVEDLDRTGPLSM <sup>TTOOSNQSPSTSHBGL</sup> EEDKHPTTSLTSSNDVDYTGRRDP	298
D	b		224 -----KDRASGGR----	231
Q	y		299 NHSEGSTTLLEGTYSHYPHHKESRTFIPVTSAKTGSFGVAIVYGDSNVVNLSLGDD	358
D	b		232 -----	231
Q	y		359 TFHDSGSHHTHGSBEDSGHSQEGGANTSGPIRTPOIPFWIIIASILLALILIAYC	418
D	b		232 -----AQTHGESFTSGHSTGSQEGGASTTSQPIRPQIPFWIIIASILLALILIAYC	284
Q	y		419 IAVNSRRCGOKKKLVINSNGAVEDRKPSGLNGEASKSQSEMVLVNKSESSEPDQMFTA	478
D	b		285 IAVNSRRCGOKKKLVINSNGAVDDKKASGLNGEASRSQEMVHLVNKSESSETPDDPMFTA	344
Q	y		479 DETRNLTQVMKIGV	493
D	b		345 DETRNLTQVMKIGV	359
<hr/>				
RESULT 6				
CD44	CANFA			
ID	CD44_CANFA	STANDARD;	PRT;	351 AA.
AC	Q28284;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	CD44 antigen precursor (Phagocytic glycoprotein I) (PGP-1) (HUTCH-I)			
DE	(Extracellular matrix receptor III) (ECM-III) (GP90 lymphocyte homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor) (Fragment).			
GN	CD44.			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
RX	NCBI_TaxId=9615;			
RL	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Beagle; TISSUE=Thymus;			
RX	MEDLINE=94250687; PubMed=7514890;			
RA	Milde K.F., Alejandro R., Mintz D.H., Pastori R.L.;			
RT	"Molecular cloning of the canine CD44 antigen cDNA."			
RL	Biochim. Biophys. Acta 1218:112-114(1994).			
CC	-1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to mucosal high endothelial venule and to types I and VI collagen. Probably involved in matrix adhesion, lymphocyte activation and lymph node homing.			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- TISSUE SPECIFICITY: Lymph nodes.			
CC	-1- PM: Extensively modified including N- and O-linked glycosylation, addition of the glycosaminoglycan chondroitin sulfate, of sulfate, of phosphate to cytoplasmic domain serine residues (By similarity).			
CC	-1- SIMILARITY: Contains 1 link domain.			
<hr/>				
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its				

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: Z27115; CAAB1630.1; -.  
 DR PIR; S45305; S45305.  
 DR HSSP; P98066; 1TSG.  
 DR InterPro; IPR001231; CD44 antigen.  
 DR InterPro; IPR000538; Link.  
 DR Pfam; PF00193; Xlink; 1.  
 DR PRINTS; PR00658; CD44.  
 DR PRINTS; PR01263; LINKMODULE.  
 DR ProDom; PD000918; Link; 1.  
 DR SMART; SM00445; Link; 1.  
 DR PROSITE; PS01241; Link; 1.  
 DR Cell adhesion; Transmembrane; Glycoprotein; Phosphorylation; Receptor;  
 KW Proteoglycan; Signal; Alternative splicing;  
 KW Pyrolydione carboxylic acid.  
 FT NON\_TER 1  
 FT SIGNAL 1  
 FT CHAIN 13  
 FT 14 >351 POTENTIAL.  
 FT 14 >351 CD44 ANTIGEN.  
 FT 14 263 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 264 284 POTENTIAL.  
 FT DOMAIN 285 >351 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 41 112 LINK.  
 FT DOMAIN 143 151 ARG/LYS-RICH (BASIC).  
 FT DOMAIN 218 263 STEM.  
 FT MOD\_RES 14 PYROLYDIONE CARBOXYLIC ACID (BY  
 FT DISULFID 46 111 SIMILARITY).  
 FT DISULFID 70 90 BY SIMILARITY.  
 FT MOD\_RES 286 286 BY SIMILARITY.  
 FT PHOSPHORYLATION (BY PKC) (PARTIAL) (BY  
 FT SIMILARITY).  
 FT MOD\_RES 320 320 PHOSPHORYLATION (PARTIAL) (BY  
 FT SIMILARITY).  
 FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 50 50 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON\_TER 351  
 SQ SEQUENCE 351 AA; 38066 MW; E73387E70E20C0E0 CRC64;  
 Query Match 56.4%; Score 1470; DB 1; Length 351;  
 Best Local Similarity 61.7%; Pred. No. 1.2e-84;  
 Matches 297; Conservative 21; Mismatches 31; Indels 132; Gaps 3;

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 QY 368 TTGHSBDSGSHSGSGAGANTTSGPIRTPOIPEMILLIILALALILAVCIANVSRRC 427  
 DB 231 TTGHSBAGHSSSGSGAGANTTSGPIRTPOIPEMILLIILALALILAVCIANVSRRC 230  
 QY 428 GQKKGLVINGANGAVBERKSSGANGERSQGVNHLVNNSSSTPQPFMTADETRLQNV 487  
 DB 291 GQKKGLVINGANGAVBERKSSGANGERSQGVNHLVNNSSSTPQPFMTADETRLQNV 350  
 QY 488 D 488  
 DB 351 D 351  
 RESULT 7  
 CD44\_BOVIN STANDARD; PRT; 366 AA.  
 ID CD44\_BOVIN  
 AC Q29423;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE CD44 antigen precursor (Phagocytic glycoprotein I) (GPC-1) (HUTCH-I)  
 DE (Extracellular matrix receptor-III) (ECMR-III) (GP90 lymphocyte  
 DE homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor).  
 GN CD44.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92017904; PubMed=1922105;  
 RA Bosworth B.T., St John T., Gallatin W.M., Harp J.A.;  
 RT "Sequence of the bovine CD44 cDNA: comparison with human and mouse  
 RT sequences.";  
 RL Mol. Immunol. 28:1131-1135(1991).  
 CC -1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to  
 CC mucosal high endothelial venule and to types I and VI collagen.  
 CC Probably involved in matrix adhesion, lymphocyte activation and  
 CC lymph node homing.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Mesenteric lymph node and liver, not in heart.  
 CC -1- PTM: Extensively modified including N- and O-linked glycosylation,  
 CC addition of the glycosaminoglycan chondroitin sulfate, of sulfate,  
 CC of phosphate to cytoplasmic domain serine residues (By  
 CC similarity).  
 CC -1- SIMILARITY: Contains 1 link domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: X62881; CAA44675.1; -.  
 DR EMBL: S63418; AAB20016.1; -.  
 DR PIR; A53286; A53286.  
 DR HSSP; P98066; 1TSG.  
 DR InterPro; IPR001231; CD44 antigen.  
 DR InterPro; IPR000538; Link.  
 DR Pfam; PF00193; Xlink; 1.  
 DR PRINTS; PR00658; CD44.  
 DR PRINTS; PR01263; LINKMODULE.  
 DR ProDom; PD000918; Link; 1.  
 DR SMART; SM00445; Link; 1.  
 DR PROSITE; PS01241; Link; 1.  
 DR Cell adhesion; Transmembrane; Glycoprotein; Phosphorylation; Receptor;  
 KW Proteoglycan; Signal; Alternative splicing;  
 KW Pyrolydione carboxylic acid.  
 FT SIGNAL 1  
 FT CHAIN 21 366 POTENTIAL.  
 FT CD44 ANTIGEN.

```

FT  DOMAIN  21  273  EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM  274  294  POTENTIAL.
FT  DOMAIN  295  366  CYTOPLASMIC (POTENTIAL).
FT  DOMAIN  448  119  LINK.
FT  DOMAIN  150  158  ARG/LYS-RICH (BASIC).
FT  DOMAIN  229  273  STEM.
FT  MOD_RES  21  21  PYROGLUTAMINE CARBOXYLIC ACID (BY
FT  SIMILARITY).
FT  DISULFID  53  118  BY SIMILARITY.
FT  DISULFID  77  97  BY SIMILARITY.
FT  MOD_RES  296  296  PHOSPHORYLATION (BY PKC) (PARTIAL) (BY
FT  SIMILARITY).
FT  MOD_RES  330  330  PHOSPHORYLATION (PARTIAL) (BY
FT  SIMILARITY).
FT  CARBOHYD  25  25  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  57  57  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  100  100  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  110  110  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  120  120  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  222  222  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  260  260  N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ  SEQUENCE  366 AA; 40001 MW; 438A5A1E631E02B4 CRC64;

```

```

Query Match 56.3%; Score 1469.5; DB 1; Length 366;
Best Local Similarity 61.0%; Pred. No. 1.4e-84;
Matches 305; Conservative 22; Mismatches 32; Indels 14; Gaps 8;

```

```

QY  1 MDKFWHAAMGLCLVPLSLAQLDNLNCRFAGVFEYKNGRYSISRTAADLCKAFNSTL 60
DB  1 MDTFWRAAMGLCLVPLSLAQLDNLNCRFAGVFEYKNGRYSISRTAADLCKAFNSTL 60
QY  61 PTAAQWEKALISGFECRCYGFIEGHVIVIRIHNSICANNNGVYILNLSQDYTCGN 120
DB  61 PTAAQWEKALISGFECRCYGFIEGHVIVIRIHNSICANNNGVYILNLSQDYTCGN 120
QY  121 ASAPREDCTSVTLDPNAPDGPITITIVNRDGRVYQGEYRTNEDI-----YPSNPTD 175
DB  121 ASAPREDCTSVTLDPNAPDGPITITIVNRDGRVYQGEYRTNEDI-----YPSNPTD 175
QY  121 ASAPREDCTSVTLDPNAPDGPITITIVNRDGRVYQGEYRTNEDI-----YPSNPTD 175
DB  121 ASAPREDCTSVTLDPNAPDGPITITIVNRDGRVYQGEYRTNEDI-----YPSNPTD 175
QY  176 DVSSSSSSSRSTSGY-LEFY-FTYVHPIDPDSFMTDSTRIPANMSSSTLQ 233
DB  176 DVSSSSSSSRSTSGY-LEFY-FTYVHPIDPDSFMTDSTRIPANMSSSTLQ 233
QY  181 DESSSSSPSR-STSGYSTFPHLPTVPSRPR-RPW-----SQRENT----- 223
DB  181 DESSSSSPSR-STSGYSTFPHLPTVPSRPR-RPW-----SQRENT----- 223
QY  234 PTANPTVGLVEDLDRTPGSMITQOSNQSFSYSHGLEDKDHPTSTLSSNNDVTG 293
DB  234 PTANPTVGLVEDLDRTPGSMITQOSNQSFSYSHGLEDKDHPTSTLSSNNDVTG 293
QY  224 -----SDTRDYSSHD----- 234
DB  224 -----SDTRDYSSHD----- 234
QY  294 GRADPNHSGSTLLRGYTSHPYTKESRTFIPVTSAGTGSFGVAVTVGDSNSVNSL 353
DB  294 GRADPNHSGSTLLRGYTSHPYTKESRTFIPVTSAGTGSFGVAVTVGDSNSVNSL 353
QY  235 ----- 234
DB  235 ----- 234
QY  354 SGQDTFHPGSGHTTHGSEDSHSGOEGANTTGPTRFQPEWTLIIILASLALAL 413
DB  354 SGQDTFHPGSGHTTHGSEDSHSGOEGANTTGPTRFQPEWTLIIILASLALAL 413
QY  235 -----PSGRSYTHASASGHSSGSEHGANTTSGMRKQIDEWIIILASLALAL 286
DB  235 -----PSGRSYTHASASGHSSGSEHGANTTSGMRKQIDEWIIILASLALAL 286
QY  414 ILAVCIANRRRCGQKKLVINSNGAVEDRKPSGLNGAASGQGMVHVNKESSETPD 473
DB  414 ILAVCIANRRRCGQKKLVINSNGAVEDRKPSGLNGAASGQGMVHVNKESSETPD 473
QY  287 ILAVCIANRRRCGQKKLVINSNGAVEDRKPSGLNGAASGQGMVHVNKESSETPD 346
DB  287 ILAVCIANRRRCGQKKLVINSNGAVEDRKPSGLNGAASGQGMVHVNKESSETPD 346
QY  474 QPMTADETRLQNVDMKIGV 493
DB  474 QPMTADETRLQNVDMKIGV 493
QY  347 QPMTADETRLQNVDMKIGV 366
DB  347 QPMTADETRLQNVDMKIGV 366

```

```

DE  homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor)
DE  (LY-24).
OS  CD44.
OS  Rattus norvegicus (Rat).
CC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC  STRAIN=BDIX; TISSUE=Pancreas;
RX  MEDLINE=91191552; PubMed=1707342;
RA  Guenther U., Hofmann M., Rudy W., Reber S., Zoeller M.,
RA  Hausmann I., Matzku S., Wenzel A., Ponta H., Herrlich P.;
RT  "A new variant of glycoprotein CD44 confers metastatic potential to
RT  rat carcinoma cells."
RL  Cell 65:13-24 (1991).
RN  [2]
RP  SEQUENCE FROM N.A. (ISOFORM 1).
RA  Stevens J.W., Midura R.J.;
RL  Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to
CC  mucosal high endothelial venule and to types I and VI collagen.
CC  Probably involved in matrix adhesion, lymphocyte activation and
CC  lymph node homing.
CC  -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC  -1- ALTERNATIVE PRODUCTS:
CC  Event=Alternative splicing; Named isoforms=2;
CC  Comment=Additional isoforms seem to exist;
CC  Name=2; Synonyms=Long, Meta-1;
CC  IsoId=P26051-1; Sequence=Displayed;
CC  Name=1; Synonyms=Short;
CC  IsoId=P26051-2; Sequence=VSP_005330;
CC  -1- PPM: N-glycosylated (By similarity).
CC  -1- PPM: O-glycosylated; contains chondroitin sulfate glycans which
CC  can be more or less sulfated (By similarity).
CC  -1- PPM: Phosphorylated; activation of PKC results in the
CC  dephosphorylation of Ser-467 (constitutive phosphorylation site),
CC  and the phosphorylation of Ser-433 (By similarity).
CC  -1- SIMILARITY: Contains 1 link domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M61875; AAAS3532.1; -
DR  EMBL; M61874; AAAS3534.1; -
DR  EMBL; U52179; AAA97915.1; -
DR  EMBL; U46957; AAA92920.1; -
DR  PIR; B38745; B38745.
DR  HSSP; P98066; ITSG.
DR  InterPro; IPR001231; CD44_antigen.
DR  InterPro; IPR000538; Link.
DR  Pfam; PF00193; Xlink; 1.
DR  PRINTS; PR00658; CD44.
DR  PRINTS; PR01265; LINKMODULE.
DR  ProDom; PD000918; Link; 1.
DR  SMART; SM00445; LINK; 1.
DR  PROSITE; PS01241; LINK; 1.
KM  Cell adhesion; Transmembrane; Glycoprotein; Phosphorylation; Receptor;
KM  Proteoglycan; Signal; Alternative splicing;
KW  Pyroglutamate carboxylic acid.
FT  SIGNAL  1  21
FT  CHAIN  22  503
FT  DOMAIN  22  410
FT  TRANSMEM  411  431
FT  DOMAIN  432  503
FT  DOMAIN  51  123
FT  DOMAIN  154  162
FT  DOMAIN  228  410
FT  STEM.

```

```

RESULT 8
CD44_RAT  STANDARD;  PRT;  503 AA.
AC  P26051; 099021;
DT  01-MAY-1992 (Rel. 22, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  15-MAR-2004 (Rel. 43, Last annotation update)
DE  CD44 antigen precursor (Phagocytic glycoprotein I) (Pgp-I) (HITCH-I)
DE  (Extracellular matrix receptor-III) (ECMR-III) (Gp90 lymphocyte

```

```

FT MOD_RES 22 22 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT DISULFID 56 122 BY SIMILARITY.
FT DISULFID 80 100 BY SIMILARITY.
FT MOD_RES 433 433 PHOSPHORYLATION (BY PKC) (PARTIAL) (BY
FT MOD_RES 467 467 SIMILARITY).
FT MOD_RES 467 467 PHOSPHORYLATION (PARTIAL) (BY
FT CARBOHYD 28 28 SIMILARITY).
FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 224 385 IATTPVSAHTKQNRQWNPPIHNSPEVLLQTTTRMDID
FT RNSTSAGENWTOEPQPPNNHRYODEESTPHATSTWADP
FT NSTTEAATQKEKPFENWQKNPPSPDSHVSHTGTTASA
FT HNNHPSQRMTOQSOEDSWTDFEDPSHMQGQHQTESK
FT -> SDGSSMDPRGDFVTYHSGEILA (in isoform
1)
FT CONFLICT 74 74 /Frtid=VSP 005330.
FT SEQUENCE 503 AA; 55945 MM; FB489D009BD4E822 CRC64;
R -> S (IN REF. 2).
Query Match 54.6%; Score 1425; DB 1; Length 503;
Best Local Similarity 58.7%; Pred. No. 1.2e-81;
Matches 315; Conservative 36; Mismatches 108; Indels 78; Gaps 16;
QY 1 MDKFWHIAANG-LCLVPLSLA--QIDINITCRAGVYFHYEKNRGRYSISTEADADCKAN 57
DB 1 MDKWHHTAWGLCLLQSLAQOQIDINITCRAGVYFHYEKNRGRYSISTEADADCEAN 60
QY 58 STLPYMAQMEKALSIGETCRGYGIEGHAVIPIPHNSICAANNVTYI-LTSNMSQYDT 116
DB 61 TLPLTMAQMEKALAKGETCRGYGIEGHAVIPIPHRAICANNVTYIILASNSHVDY 120
QY 117 YCFNAPAPPEEDCTSVTDLFNAPDPTITITVNRDGRYQKEVYTNPEDIYPSPTD 176
DB 121 YCFNAPAPPEEDCTSVTDLFNAPDPTITITVNRDGRYQKEVYTNPEDIYPSPTD 180
QY 177 DVSSGSSSESTSGYIYFT-ESTVAPRDEE-----SPITDST---D 217
DB 181 DVSSGSSSESTSGYIYFT-ESTVAPRDEE-----SPITDST---D 217
QY 218 RIPATNMDSSHITLQPTANPNTGLVEDLR--TGPISMTQSSQNSQSFSTSHGLEEDK 275
DB 240 RTQWNPPIHNSPEVLLQTTTR---MTDIDRNSGSAHGENWTOEPQPPNNHRYODEEST 294
QY 276 DHPITSLTSSNRDVG-----GRDDPHSGSGTTLLEGYT--SHYPTKE 320
DB 295 PHATSTWADPNSTTEEAATQKEKPFENWQKNPPSPDS-HVTEGTTASAHNNHPSQ 353
QY 321 SRTEIPYTSAGTSFGVAVYAVVGDNSNVRSLSGDDTF---HPGSGHTTHGSESDG 376
DB 354 RMT-----TQSGEDVSWT-----DFPDPIHSHFGQGHQT---ESKX 386
QY 377 HSHGSOEGANTTSGPIRTPOIPEWLIILASLALAILAVCAVNSRRRCGGKKLVIN 436
DB 387 HSSNQDSGVTTTGGPARPOIPEWLIILASLALAILAVCAVNSRRRCGGKKLVIN 446
QY 437 SGNQAVEDRKPSGLNGEASKSQEMVHLVKNKESSETPQFMTADETRNLQNVDKIGV 493
DB 447 SGNQAVEDRKPSGLNGEASKSQEMVHLVKNKESSETPQFMTADETRNLQNVDKIGV 503

```

```

DE CD44 antigen precursor (phagocytic glycoprotein I) (PGE-1) (HUTCH-I)
DE (Extracellular matrix receptor-III) (ECMR-III) (Gp90 lymphocyte
DE homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor).
DE CD44.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
CX NCBI_Taxid=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90304889; Pubmed=1694723;
RA Aruffo A., Stamenkovic I., Melnick M., Underhill C.B., Seed B.;
RT "CD44 is the principal cell surface receptor for hyaluronate.";
RL Cell 61:1303-1313(1990).
CC -1- FUNCTION: Main cell surface receptor for hyaluronate. Adhesion to
CC mucosal high endothelial venules and to types I and VI collagen.
CC Probably involved in matrix adhesion, lymphocyte activation and
CC lymph node homing.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- INDUCTION: By EBV.
CC -1- PM: Extensively modified including N- and O-linked glycosylation,
CC addition of the glycosaminoglycan chondroitin sulfate, of sulfate,
CC of phosphate to cytoplasmic domain serine residues.
CC -1- SIMILARITY: Contains 1 link domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M33827; AAA36967.1; -.
DR PIR: A35616; A35616.
DR HSSP: P98066; ITSG.
DR InterPro: IPR001231; CD44 antigen.
DR InterPro: IPR005038; Link.
DR Pfam: PF00193; XLink; 1.
DR PRINTS: PR00658; CD44.
DR PRINTS: PR01265; LINKMODULE.
DR ProDom: PD00918; Link; 1.
DR SMART: SMO0445; Link; 1.
DR PROSITE: PS01241; LINK; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; Phosphorylation; Receptor;
KW Proteoglycan; Signal; Alternative splicing;
KW Pyrrrolidone carboxylic acid.
FT SIGNAL 1 22
FT CHAIN 23 362
FT DOMAIN 23 269
FT TRASMEN 270 290
FT DOMAIN 291 362
FT DOMAIN 50 121
FT DOMAIN 152 160
FT DOMAIN 226 269
FT MOD_RES 23 23
FT DISULFID 55 120
FT DISULFID 79 99
FT MOD_RES 292 292
FT MOD_RES 326 326
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 362 AA; 39775 MM; E89E834E9EC948 CRC64;

```

Query Match 53.6%; Score 1398.5; DB 1; Length 362;  
 Best Local Similarity 58.4%; Pred. No. 3.5e-80;  
 Matches 291; Conservative 26; Mismatches 40; Indels 141; Gaps 7;

QY 1 MKPFWHAMGLCYPLSLA--QIDILNTRCAGVFEHKKNGRYISRTPEADCKAPNS 58  
 DB 1 MKPFWHAMGLCYPLSLA--QIDILNTRCAGVFEHKKNGRYISRTPEADCKAPNS 60  
 QY 59 TLPTAOMKALSIGETCRYPFIEGHVPIPHNS:CAANNVTGVTLLTNTSGDYTC 118  
 DB 61 TLPTAOMKALSIGETCRYPFIEGHVPIPHNS:CAANNVTGVTLLTNTSHYDYTC 120  
 QY 119 FNASAPREDCTSVNDLPNAPDGPITITVNDGRTYVCKGRTNPDYPSNPTDDV 178  
 DB 121 FNASAPLEDCTSVNDLPNAPDGPITITVNDGRTYVCKGRTNPDYPSNPTDDV 180  
 QY 179 SSGSSRSSTSGYTFYTF-STVHPIDPDSFWITDSTRIPATNMSHSTTLQPTAN 237  
 DB 181 SSGSSRSSTSGYTFYTF-STVHPIDPDSFWITDSTRIPATNMSHSTTLQPTAN 218  
 QY 238 PNTGLVEDLDRTGPIASMTTQNSQSFSTSHGLEDHDPPTSTLTSNRDVTGERD 297  
 DB 219 -----MATR----- 222  
 QY 298 PNHSEGSTLLBEGYTSHPYTESRTFIPVTSKATGSGFVATVVGDSNVNRSLSGDQ 357  
 DB 223 -----DQ 224  
 QY 358 D-TFHPGSGSH-TGHSBDSGSHSQSGGANTTSGPITPDIPEWLLIASLALAIL 415  
 DB 225 DSGMDPRGNSLFTVTDSSKLTSHSGSGQDGLNSTSPGKPRVPMWLVLSLALAIL 284  
 QY 416 AVCIANVSRRCQCKKLVINSNGAVEDRKPSGNGEASQCEWLVNKSSETPDQF 475  
 DB 285 AVCIANVSRRCQCKKLVINSNGAVEDRKPSGNGEASQCEWLVNKSSETPDQF 344  
 QY 476 MTADEFRLQNVDMKIGV 493  
 DB 345 MTADEFRLQNVDMKIGV 362

RESULT 10  
 SERI\_BOMMO STANDARD; PRT; 389 AA.

AC P07856;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sericin precursor (Silk gum protein).  
 OS Bombyx mori (Silk moth).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;  
 OC Bombycidae; Bombyx.  
 OX NCBI\_TaxID=7091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Middle silk gland;  
 RA MEDLINE=6306283; PubMed=6294094;  
 RA Okamoto H., Ishikawa S., Suzuki Y.,  
 RT "Structural analysis of sericin genes. Homologies with fibroin gene  
 RT in the 5' flanking nucleotide sequences.";  
 RT J. Biol. Chem. 257.15192-15199 (1982).  
 RN [2]  
 RP SEQUENCE OF 317-354 FROM N.A.  
 RX MEDLINE=67076763; PubMed=3024742;  
 RA Michaille J.-C., Garel A.,  
 RT "A single gene produces multiple sericin messenger RNAs in the silk  
 RT gland of Bombyx mori.";  
 RT Biochimie 68:1165-1173 (1986).  
 CC -1- FUNCTION: Provides the silk fibroin thread with a sticky coating.  
 CC Acts as a cement by sticking silk threads together.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Produced exclusively in the middle (MSG)

CC section of silk glands.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; J01029; NOT ANNOTATED CDS.  
 CC EMBL; J01030; NOT ANNOTATED CDS.  
 CC EMBL; J01031; NOT ANNOTATED CDS.  
 CC EMBL; J01032; NOT ANNOTATED CDS.  
 CC EMBL; J01033; NOT ANNOTATED CDS.  
 CC EMBL; J01034; NOT ANNOTATED CDS.  
 CC EMBL; M26101; AAA27843.1; --  
 CC EMBL; M26102; AAA27844.1; --  
 CC EMBL; Signal.  
 CC SIGNAL 1 15  
 CC CHAIN 16 389  
 CC FT SERICIN.  
 CC SEQUENCE 389 AA; 38820 MW; BA605BC0305EAF19 CRC64;

Query Match 6.8%; Score 177; DB 1; Length 389;  
 Best Local Similarity 23.8%; Pred. No. 0.00036;  
 Matches 101; Conservative 55; Mismatches 131; Indels 138; Gaps 18;

QY 38 KNGRYSIRTEADLCFAFNSTLPTAOMEKAL-----SIGFET 76  
 DB 32 KNRKTAASSESYLND-NDISAGAHRAKSYGQODSKYTSRGEVYSGRQNYKD 90  
 QY 77 CRYPFIEGHVPIPHNS:CAANNVTGVTLLTNTSGDY-TYFNASAPREDCTSVTL 135  
 DB 91 SKQALISGCT---KSSNVQSDSKSAGSSSSRSQESAYSSSSSTEESSSSSRA 147  
 QY 136 PNAFDGPITITVNDGRTYVCKGRTNPDYPSNPTDDVS--GSSSER----- 186  
 DB 148 ASSTDAS-SNTDSNSNAGSSTGGRTTYG---YSSNRSRGVSSTGSSSNTDSNSNAG 203  
 QY 187 SSTSGGYTFYTFSTVHPIDPDSFWITDSTRIPATNMSHSTTLQPTANPTGLVEDL 246  
 DB 204 SSTSGGSSSTGYGVS-----SNRDSGVSST----- 226  
 QY 247 DRGPIASMTTQNSQSFSTSHGLEDHDPPTSTLTSNR-----NDVTG 293  
 DB 227 --TSSSNTDSNSNAGSSTGGSSSHEDSKSDENVSTTSSSNTDSNSNAGSSTG 284  
 QY 294 GRR-----DPN-HSEGSTLLBEGYTSHPYTESRTFIPVTS 330  
 DB 285 GRATYGVSSNRDGSVSGTSSSNTDSNSNAGSST--SGGSSSTYGVSSNSR--DGSVS 339  
 QY 331 KTGFQYTAATVGDNSNNVRSLSGDDTPHPGSGSTTHGSSDGHSHSQSGGANTTS 390  
 DB 340 STGSSSNT-----DSNSN-----SAGSSTSGGSSSTYGVSSNSHDSVSGT 380  
 QY 391 GPIRT 395  
 DB 381 SSSNT 385

RESULT 11  
 TSG6\_RABIT STANDARD; PRT; 276 AA.

AC P98065;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Tumor necrosis factor-inducible protein TSG-6 precursor (TNF-  
 DE stimulated gene 6 protein) (Hyaluronate-binding protein P84).  
 GN TNFIP6 OR TSG6 OR P84.  
 GN Oryctolagus cuniculus (Rabbit).  
 OS Oryctolagus cuniculus; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Eukaryota; Metazoa; Chordata; Lagomorpha; Leporidae; Oryctolagus;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

DR	PIR: A48055; A47290.
DR	HSSP: P98066; ITSG.
DR	InterPro: IPR000959; CUB.
DR	InterPro: IPR000538; Link.
DR	Pfam: PF00431; CUB; 1.
DR	Pfam: PF00193; XLink; 1.
DR	PRINTS: PR01265; LINKMODLE.
DR	ProDom: PD000918; Link; 1.
DR	SMART: SM00042; CUB; 1.
DR	SMART: SM00445; LINK; 1.
DR	PROSITE: PS01180; CUB; 1.
DR	PROSITE: PS01241; Link; 1.
KM	cell adhesion; Signal; Glycoprotein.
FT	SIGNAL
FT	CHAIN
FT	20
FT	276
FT	15
FT	POTENTIAL.
FT	TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN
FT	TSF-6.
FT	CUB.
FT	LINK.
FT	DOMAIN
FT	53
FT	128
FT	135
FT	247
FT	DISEUFD
FT	58
FT	127
FT	DISEUFD
FT	82
FT	103
FT	DISEUFD
FT	135
FT	161
FT	DISEUFD
FT	188
FT	210
FT	CARBOHYD
FT	118
FT	118
FT	258
FT	258
FT	CARBOHYD
FT	258
FT	276 AA;
FT	31081 MM;
FT	3BDCSD9A24A2F75A CRCS4;
FT	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	N-LINKED (GLCNAC .) (POTENTIAL).
FT	SEQUENCE

[illegible]

```

Cy      164 N--PEDIYPSNPDDDDSSGSS-----SSRPSGSGYFYFTSTYHP 203
Db      212 DELPEDI-----ISTGNVNTLAKFLSDASVYTAGGFQI-KVYVDP 249

RESULT 12
PGCA RAT STANDARD; PRT; 2124 AA.
AC P07897;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
OS AGC1 OR AGC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86087070; PubMed=3693370;
RA Doerge K., Saeki M., Horigan E., Haesell J.R., Yamada Y.;
RT "Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones."
RL J. Biol. Chem. 262:1757-1767(1987).
RN [2]
RP REVISION TO 698.
RX Doerge K., Saeki M., Horigan E., Haesell J.R., Yamada Y.;
RA J. Biol. Chem. 263:10040-10040(1988).
RN [3]
RP SEQUENCE OF 1856-2124 FROM N.A.
RX MEDLINE=86250698; PubMed=2424893;
RA Doerge K., Fernandez P., Haesell J.R., Saeki M., Yamada Y.;
RT "Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan."
RL J. Biol. Chem. 261:8108-8111(1986).
CC -1- FUNCTION: This proteoglycan is a major component of extracellular matrix of cartilaginous tissues. A major function of this protein is to resist compression in cartilage. It binds avidly to hyaluronic acid via an amino-terminal globular region. May play a regulatory role in the matrix assembly of the cartilage.
CC -1- SUBUNIT: Interacts with PRN1 (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
CC -1- DOMAIN: Two globular domains, G1 and G2, comprise the amino terminus of the proteoglycan, while another globular region, G3, makes up the COOH terminus. G1 contains link domains and thus consists of three disulfide-bonded loop structures designated as the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS) and the chondroitin sulfate (CS) attachment domains lie between G2 and G3.
CC -1- PTM: Contains mostly chondroitin sulfate, but also keratan sulfate chains. N-linked and O-linked oligosaccharides.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 4 link domains.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33518; AAA41836.1; -
DR EMBL; J03485; AAA21000.1; ALT_SEQ.
DR F01; A92623; A28452.
DR HSP; P98066; ITSG.
DR InterPro; IPR002353; AntiIfzetae11

```

```

DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR InterPro: IPR001304; Lectin_C.
DR InterPro: IPR000538; Link.
DR InterPro: IPR003324; SGXSG.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00047; Ig_1.
DR Pfam: PF00059; Lectin_C_1.
DR Pfam: PF02339; SGXSG_55.
DR Pfam: PF00084; Sushi_1.
DR Pfam: PF00193; Xlink; 4.
DR PRINTS: PR00356; ANTIPREZEIT.
DR PRINTS: PR01265; LINKMODULE.
DR PRODOM: PD000918; Link; 4.
DR SMART: SM00032; CCP; 1.
DR SMART: SM00034; CLECT; 1.
DR SMART: SM00406; IGV; 1.
DR SMART: SM00445; LINK; 4.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
DR PROSITE: PS50835; IG_LIKE; 1.
DR PROSITE: PS00290; IG_MHC; 1.
DR PROSITE: PS01241; LINK; 4.
KW Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; Repeat;
KW Immunoglobulin domain.
FT SIGNAL 1 19
FT CHAIN 20 2124
FT DOMAIN 34 147
FT DOMAIN 170 247
FT DOMAIN 268 349
FT DOMAIN 504 581
FT DOMAIN 602 683
FT DOMAIN 1910 2036
FT DOMAIN 2040 2098
FT DOMAIN 48 140
FT DOMAIN 152 247
FT DOMAIN 253 349
FT DOMAIN 486 580
FT DOMAIN 587 682
FT DOMAIN 685 798
FT DOMAIN 801 1226
FT DOMAIN 1227 1909
FT DOMAIN 1910 2124
FT DISULFID 51 133
FT DISULFID 175 246
FT DISULFID 199 220
FT DISULFID 273 348
FT DISULFID 297 318
FT DISULFID 509 580
FT DISULFID 533 554
FT DISULFID 607 682
FT DISULFID 631 652
FT DISULFID 1914 1925
FT DISULFID 1942 2034
FT DISULFID 2010 2026
FT DISULFID 2041 2084
FT DISULFID 2070 2097
FT CARBOHYD 126 126
FT CARBOHYD 239 239
FT CARBOHYD 333 333
FT CARBOHYD 387 387
FT CARBOHYD 611 611
FT CARBOHYD 667 667
FT CARBOHYD 1842 1842
SQ SEQUENCE 2124 AA; 22117 MW; E30BBE61593A34B1 CRC64;

```

```

Query Match 6.4%; Score 168; DB 1; Length 2124;
Best Local Similarity 20.0%; Pred. No. 0.012;
Matches 99; Conservative 55; Mismatches 173; Indels 168; Gaps 18;
48 EAADLCAPNSTPTMAQMEKALSIGETCRVGFII-EGHVVIPRIHPNSICAMNTG---103

```

```

Db 602 EAQAFCAQNAATLASTGQLVAAMSQGLDKCYAGMLADGTLRYPIVNPBACGDKXPQVRT 661
Oy 104 VYIITSNT-----SQDYTCFNA----- 121
Db 662 VYLYENQGTGLPDLKSHKACFCFRGVSVSPGGTPTSPDIEDMIYTRVPGVDAVLEP 721
Oy 122 -----SAPPEEDCT-----SVT 133
Db 722 ETEVVPYFTEPEKQTEMEBPATYPVGNPLPGIPPTMLPVPAABEHTESPASQEPAS 781
Oy 134 DLPNAPDPTITITIVNRDGT-----RYVKGERTNPED 167
Db 782 QVPSTSEPTPLSLAPSSGTELPSSGDTSGADPLSGDTFTSGRLDSSGEPGSGSES 841
Oy 168 IYSPNPPTDD-----VSSGSSERSSTGCVIYFTESYVHPIDEDSPW-ITDSTDRIPA 221
Db 842 GLPBGDLSSGLGPTVSSGLPVESGSASG-----DGEIPWASTPTVDRLP 887
Oy 222 TNMDSHSTLTQPTAN-----PNTGLVEDLDRTPGLSKTTQSSQSFSTHGLEBKCH 277
Db 888 GGEILGSSASASGTDLSGLPSSGELITETASGTEELISGLPSSGDDLETSTGIDGASVL 947
Oy 278 PTTG---TLTSSNNDVTGGRDPNHEGSTTILEGX-----TSHYPTKES---RT 323
Db 948 PTGRGGLFETASGVEDLSG---LPSSGEGSETSTGIEDLSVLPFGSPPTASGVDDLS 1004
Oy 324 FIPV--TSANTGSPGVTAVT-VGDSNSNVNRSLSGDD-THFGSGSH---THGSESDG 376
Db 1005 GLPBGGLSTETASGVADVTLQPTERGLTETASGIEDIVLPTGHEMLTETASGVEDVS 1064
Oy 377 HSHSGREGANTSG 391
Db 1065 GLPBGKGLTETASG 1079

RESULT 13
TSG6_MOUSE STANDARD; PRT; 275 AA.
ID TSG6_MOUSE
AC 008859
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tumor necrosis factor-inducible protein TSG-6 precursor (TNF-
stimulated gene 6 protein).
DE TNFAIP6 OR TNFIP6 OR TSG6.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A. AND DEVELOPMENTAL STAGE.
STRAIN=CD-1; TISSUE=Cumulus cell, Embryo, and Oocyte;
MEDLINE=98087423; PubMed=9427551;
RA Fuetloep C., Kamath R.V., Li Y., Otto J.M., Salusri A., Olsen B.R.,
Glant T.T., Hascaill V.C.;
"Coding sequence, exon-intron structure and chromosomal localization
of murine TNF-stimulated gene 6 that is specifically expressed by
RT expanding cumulus cell-oocyte complexes.";
RT Gene 202:95-102(1997).
RL FUNCTION: Possibly involved in cell-cell and cell-matrix
interactions during inflammation and tumorigenesis (By
similarity).
-!- DEVELOPMENTAL STAGE: Expressed in cumulus cell-oocyte complexes
during expansion in vivo.
-!- SIMILARITY: Contains 1 link domain.
-!- SIMILARITY: Contains 1 CUB domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/)

```

CC or send an email to [license@ib-sib.ch](mailto:license@ib-sib.ch).

CC EMBL; U83903; AAC3527.1; -

DR PIR; J06506; J06506.

DR HESP; P98066; TSG.

DR MGI; M1195266; Tnfalp6.

DR InterPro; IPR000859; CUB.

DR InterPro; IPR000538; Link.

DR Pfam; PF00431; CUB; 1.

DR Pfam; PF00193; XLink; 1.

DR PRINTS; PR01265; LINKMODULE.

DR ProDom; PD000918; Link; 1.

DR SMART; SM00042; CUB; 1.

DR SMART; SM00445; Link; 1.

DR PROSITE; PS01180; CUB; 1.

DR PROSITE; PS01241; Link; 1.

DR Cell adhesion; Signal; Glycoprotein.

KW SIGNAL 1 17 POTENTIAL.

FT CHAIN 18 275 TUMOR NECROSIS FACTOR-INDUCIBLE PROTEIN

FT DOMAIN 35 128 TSG-6.

FT DOMAIN 135 247 CUB.

FT DISULFID 58 127 BY SIMILARITY.

FT DISULFID 82 103 BY SIMILARITY.

FT DISULFID 135 161 BY SIMILARITY.

FT DISULFID 188 210 BY SIMILARITY.

FT CARBOHYD 118 118 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 258 258 N-LINKED (GLCNAC... ) (POTENTIAL).

SQ SEQUENCE 275 AA; 30924 MW; 1CD247228208BF9 CRC64;

Query Match 6.2%; Score 163; DB 1; Length 275;

Best Local Similarity 26.1%; Pred. No. 0.0017;

Matches 63; Conservative 34; Mismatches 76; Indels 68; Gaps 13;

CC

QY 31 AGVFWV-KNGRYSISRTADCKAFNSTLPTMAQMEKALSIGFTCYGFI-EGHYVI 88

DB 35 AGVYHRARAGRYKLTVAEKAVCEPFGGLATYKQLEARKIGFVCAAGMAYKRGVY 94

QY 89 PRHENSICANNITGV---ILTSNTSQYDTTCFNASAPPEEDCTSY-TD-----L 135

DB 95 PIYKFGPCGRKGTIGIIVGIRLRSERWDAYCYN---PAAKCGGVFTDPRIKPSGF 151

QY 136 PNAFDG-----PITTYNRDGTGTR-----YVQ-----KGEYRT 163

DB 152 PNRVDNQVCYWHIRLKGRIHSLFDPLHDPGLADYVAIIVSYDVGFRVYCG 211

QY 164 N--PEDIYPSNPITDDVSGS-----SSERSSTGGYFYFTSTVHPIDEDSPWITDS 215

DB 212 DELPEDI-----ISTGNVMTLKFSLDASYTAGGFQI-KYVTVDPASKSQANKTST 261

QY 216 T 216

DB 262 T 262

RESULT 14

TSG6 HUMAN STANDARD; PRT; 277 AA.

AC P98066; O8MW19;

DT 01-FEB-1986 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Tumor necrosis factor-inducible protein TSG-6 precursor (TNF-stimulated gene 6 protein) (Hyaluronate-binding protein).

GN TNFAIP6 OR TSG6.

OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

CC NCBI\_TaxID=9606;

OX [1]

RP SEQUENCE FROM N.A.

RC TISSUE= Fibroblast;

RX MEDLINE=92113993; PubMed=1730767;

RA Lee T.H., Wisniewski H.-G., Vilcek J.;

RT "A novel secretory tumor necrosis factor-inducible protein (TSG-6) is a member of the family of hyaluronate binding proteins, closely related to the adhesion receptor CD44.";

RT J. Cell Biol. 116:545-557(1992).

RL [2]

RP SEQUENCE FROM N.A., AND VARIANT GLN-144.

RX MEDLINE=21975206; PubMed=1854277;

RA Nentwich H.A., Mustafa Z., Rugg M.S., Marsden B.D., Cordell M.R., Mahoney D.J., Jenkins S.C., Dowling B., Fries E., Milner C.M., Loughlin J., Day A.J.;

RT "A novel allelic variant of the human TSG-6 gene encoding an amino acid difference in the CUB module. Chromosomal localization, frequency analysis, modeling, and expression.";

RT J. Biol. Chem. 277:15354-15362(2002).

RL [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Lung, and Spleen;

RX MEDLINE=22386257; PubMed=12477932;

RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D., Altschul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.J., Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.D., Hulik S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butlerfield Y.S.N., Krzywinski M.I., Skalka U., Smalins D.E., Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).

RL [4]

RP STRUCTURE BY NMR OF 36-133.

RX MEDLINE=96390850; PubMed=8797823;

RA Kohda D., Morton C.J., Parkar A.A., Hatanaka H., Inagaki F.M., Campbell I.D., Day A.J.;

RT "Solution structure of the link module: a hyaluronan-binding domain involved in extracellular matrix stability and cell migration.";

RL Cell 86:767-775(1996).

CC -1- FUNCTION: Possibly involved in cell-cell and cell-matrix interactions during inflammation and tumorigenesis.

CC -1- INDUCTION: By TNF.

CC -1- SIMILARITY: Contains 1 CUB domain.

CC -1- SIMILARITY: Contains 1 CUB domain.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements) or send an email to [license@ib-sib.ch](mailto:license@ib-sib.ch)).

CC

CC EMBL; M31165; AB00792.1; -

CC EMBL; AJ421518; CAD13434.1; -

CC EMBL; AJ419936; CAD12353.1; -

CC EMBL; BC030205; AAH30205.1; -

CC PIR; A41735; A41735.

CC PDB; 1O7B; 07-NOV-03.

CC Genew; HGNC:11898; TNFAIP6.

CC MIM; 600410;

CC GO; GO:0004895; F:cell adhesion receptor activity; TAS.

CC GO; GO:0005540; F:hyaluronic acid binding; TAS.

CC GO; GO:0007267; P:cell-cell signaling; TAS.

CC GO; GO:0006954; P:inflammatory response; TAS.



DR GO; GO:0007165; P:signal transduction; TAS.  
DR InterPro: IPR000859; CUB.  
DR InterPro: IPR000538; Link.  
DR Pfam: PF00431; CUB; 1.  
DR Pfam: PF00193; Xlink; 1.  
DR PRINTS; PR01265; LINKMODULE.  
DR Prodom; PD000918; Link; 1.  
DR SMART; SMO0042; CUB; 1.  
DR SMART; SMO0445; Link; 1.  
DR PROSITE; PS01180; CUB; 1.  
DR PROSITE; PS01241; Link; 1.  
KW Cell adhesion; Signal; Glycoprotein; Polymorphism; 3D-structure.  
FT SIGNAL 1  
FT CHAIN 20 277  
FT FT  
FT FT  
FT DOMAIN 53 128  
FT DOMAIN 135 247  
FT DISULFID 58 127  
FT DISULFID 82 103  
FT DISULFID 135 161  
FT DISULFID 188 210  
FT CARBOHYD 118 118  
FT CARBOHYD 258 258  
FT VARIANT 144 144  
FT FT  
FT STRAND 38 39  
FT HELIX 51 60  
FT TURN 71 72  
FT STRAND 85 85  
FT STRAND 88 88  
FT TURN 89 90  
FT STRAND 91 91  
FT TURN 103 106  
FT TURN 116 117  
FT STRAND 124 124  
FT STRAND 126 127  
SQ SEQUENCE 277 AA; 31231 MW; 4DB3ABE4AC52B880 CRC64;  
  
Query Match 6.2%; Score 162.5; DB 1; Length 277;  
Best local similarity 33.3%; Pred. No. 0.0019;  
Matches 40; Conservative 21; Mismatches 50; Indels 9; Gaps 5  
  
OY 29 RFAGVPHVE-KNGRYSISRTAADCKKFNGLPTMAOMEKALSIGPCTGSGFI-EGRV 86  
Db 33 RAGGVHREKRBSKYTLTAEKAKNCEFEGGHILATYKLEAARKLIGHVCAAGMAMKRV 92  
OY 87 VIFRIHPNSICAAANTGV---ILTSNTSYDTTCFNASAPPEEDCTSV-TDLPNAPDG 142  
Db 93 GYIVIVRPENCGFGKGITIDYGIRLINRBERMDACYN---PHAKECGGVTDPKRLTFKSP 149

RESULT 15  
MUC1\_MOUSE  
ID MUC1\_MOUSE STANDARD; PRT; 630 AA.  
AC Q02496;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mucin 1 precursor (Polymorphic epithelial mucin) (PEWT) (Epistatin).  
GN MUC1 OR MUC-1.  
OS Mus musculus (Mouse).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91332029; PubMed=1714452;  
RA Splicer A.P., Parry G., Patton S.J.;  
RT "Molecular cloning and analysis of the mouse homologue of the tumor-associated mucin, MUC1, reveals conservation of potential O-glycosylation sites, transmembrane, and cytoplasmic domains and a loss of minisatellite-like polymorphism.";  
RT J. Biol. Chem. 266:15099-15109(1991).

[illegible]

FT	CONFLICT	121	S (IN REF. 2).
FT <th>CONFLICT</th> <td>138 <td>AT -&gt; PA (IN REF. 2).</td> </td>	CONFLICT	138 <td>AT -&gt; PA (IN REF. 2).</td>	AT -> PA (IN REF. 2).
FT <th>CONFLICT</th> <td>139 <td>T -&gt; TT (IN REF. 2).</td> </td>	CONFLICT	139 <td>T -&gt; TT (IN REF. 2).</td>	T -> TT (IN REF. 2).
FT <th>CONFLICT</th> <td>140 <td>F -&gt; S (IN REF. 2).</td> </td>	CONFLICT	140 <td>F -&gt; S (IN REF. 2).</td>	F -> S (IN REF. 2).
FT <th>CONFLICT</th> <td>423 <td>S -&gt; D (IN REF. 2).</td> </td>	CONFLICT	423 <td>S -&gt; D (IN REF. 2).</td>	S -> D (IN REF. 2).
FT <th>CONFLICT</th> <td>506 <td>S -&gt; S (IN REF. 2).</td> </td>	CONFLICT	506 <td>S -&gt; S (IN REF. 2).</td>	S -> S (IN REF. 2).
FT <th>CONFLICT</th> <td>602 <td>Q</td> </td>	CONFLICT	602 <td>Q</td>	Q
Q <th>SEQUENCE</th> <td>630 AA:</td> <td>FEF57CIBJ3137683B CRC64</td>	SEQUENCE	630 AA:	FEF57CIBJ3137683B CRC64

FT	CONFLICT	121	121	L -> S (IN REF. 2).
FT	CONFLICT	138	139	AT -> PA (IN REF. 2).
FT	CONFLICT	140	140	AT (IN REF. 2).

ET	CONFLICT	140	140	1 - 3 (IN REF. 2).
ET	CONFLICT	433	433	1 - 3 (IN REF. 2).

ET	CONFLICT	506	506	S	D	(IN REF. 2)
F1	CONFLICT	423	423	4	3	

FT CONFLICT 602  
CONFLICT 602  
O - V S (IN BEEF : 2)

---

CONFIDENTIAL

SO SEQUENCE 630 AA: 64622 MW: FF57C1B31137C83B CH

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

Query Match 6.1%; Score 158; DB 1; Length 630;

Best Local Similarity 26.2%; Pred. No. 0.01;

Matches	84;	Conservative	42;	Mismatches	136;	Indels	58;	Gaps	15;
---------	-----	--------------	-----	------------	------	--------	-----	------	-----

```

Qy      8  IEGHVIYPRHPMSICAMNTGVILITNSITSCYDTCFMAAP----- 125
      18  LKGFILALPS-EENSIVTSSODTSSSLASTTTPVHASNSDPAIRPFGDSTSSPVQSTSSPA 76
Db      126  ---BEDCTSYDLPNAPDGEITITTVNRDGRVYQKGRYRNPEDIVPSNFTDDVASGS 182
      77  TRAPEOSTS-TAVLSGTSPPATTAAPN-SASPVAHG-----TSSPATSPLDSN 125
Qy      183  SSE-RSSTSGGYIFTFSTVHPIDPEDEPMITSDTRIPATNM--DSHSHTLOPTANP 238
      126  SSPVHSGTSSA-----ATTAPVDSTSPVHGGTSS-PATSPGPGSTSPDHSGTSSP 178
Db      239  NTGLVEDLDRTGLSMTTQOONSQSF-STSHSGLEEDKHPTSTLTSSNRNDYT-GGR 236
      179  ATRAPEBSTSTAVLSGTSPPATTAAPVDSTSSPVAHDDTSSPSTLSIEDSASPVAHGOTS 238
Qy      297  DPNHSEGSTLLEGYSHYPHTKESRFLPVTS--AKTSGFQVATVATGDSNVNNSLS 354
      239  SP-----ATSPFLNDSTSSPVHSGASIQNIKTTSDLASPPDHNGTSVTL--TSALGSA 231
Db      355  GDDDTFHPSPGSHTHGSES 374
Qy      292  PDH-----SGTSTTNSSS 306
      292  PDH-----SGTSTTNSSS 306

```

Db 18 LKGFALPS-EENSVTSSQDTSSSLASTTTPVHSSNSDPATRPDSTSSPVQSSSTSSPA 76

126 ---EEDCTSVTDLPNAFDGPITITIVNRDGTXYVQKGEYRTNPEDIYPSNPTDDDVSSGS 182

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

183 SSE--KSSISGGYLFYIFSVAPLEDEUSPMTJDSIDKLPAINM--DSSHSILQIFANF 238

DB 126 SSPVHSGISSA-----A11APVDSISSPVHGGISS-PA1SPGDSISSPLHSSISSP 1/8

226 2022-2023

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

DD 179 A I K A P E D S I S I A V L S G I S S P A T I A P V D S I S S P V A H D I S S P A I S L S E D S A S S P V A H G G I S 238

29 / DENHSGSTILLEGIISHYPIHESKRIE L PVIS--ALIGSFVIAI VGDNSNVKBSLS 334

```

220  CD - - - - - ATCDBTPOETSCBYHSSASTONTKTSTDI.ASTBDHNGTGVTT - - - - - TS8A1G8ATS 291

```

Search completed: March 8, 2004, 06:08:51  
Job time : 20 secs

Job time : 20 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2004, 06:08:16 ; Search time 45 Seconds

(without alignments)  
3456.677 Million cell updates/sec

Title: PCT-US01-51014-1

Perfect score: 2608  
Sequence: 1 MDKFWHAAWGLCLVPLSLA.....QFMADETRLQNVDMKIGV 493

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2608	100.0	493	4	Q86227
2	2490	95.5	493	4	Q96124
3	2466	94.6	470	4	Q9H5A6
4	2457.5	94.2	742	4	Q9U336
5	2348	90.0	676	4	Q9H5A7
6	2326.5	89.2	719	4	Q9H5A5
7	2189.5	84.0	514	11	Q62913
8	1837	70.4	361	4	Q86772
9	1835	70.4	361	4	Q8N694
10	1748.5	67.0	580	11	Q80X37
11	1705	65.4	338	4	Q9H5A4
12	1637.5	62.8	780	11	Q08779
13	1536.5	58.9	364	6	Q97569
14	1494	57.3	294	4	Q92493
15	1367.5	52.4	364	11	Q70509
16	1358	52.1	271	4	Q9H5A3

17	871	33.4	177	4	Q99900	Q99900 homo sapien
18	836.5	32.1	398	13	Q90ZL8	Q90L18 anas platyr
19	804.5	30.6	396	13	Q9W6S4	Q9W6S4 gallus gall
20	662	25.4	127	4	Q9P222	Q9P222 homo sapien
21	646	24.8	183	6	Q9T0U8	Q9T0U8 bos taurus
22	638	24.5	183	6	Q9T0U8	Q9T0U8 bos indicus
23	634	24.3	183	6	Q9T0U8	Q9T0U8 bos taurus
24	502	19.2	265	13	Q98SR5	Q98SR5 anas platyr
25	450	17.3	168	13	Q90ZL6	Q90ZL6 anas platyr
26	432	16.6	106	5	Q17448	Q17448 strongyloce
27	422	16.2	437	11	Q35249	Q35249 rattus norv
28	417	16.0	80	4	Q86U21	Q86U21 homo sapien
29	417	16.0	139	4	Q95370	Q95370 homo sapien
30	200	7.7	318	11	Q8BHC0	Q8BHC0 mus musculu
31	200	7.7	318	11	Q93NE4	Q93NE4 mus musculu
32	199.5	7.6	322	4	Q8TCL8	Q8TCL8 homo sapien
33	199.5	7.6	322	4	Q9UNF4	Q9UNF4 homo sapien
34	192.5	7.4	332	4	Q95Y7	Q95Y7 homo sapien
35	185.5	7.1	162	11	Q93RN2	Q93RN2 rattus norv
36	183	7.0	222	5	Q81RX6	Q81RX6 caenorhabdi
37	178	6.8	833	16	Q98Q44	Q98Q44 mycoplasma
38	172	6.6	94	11	Q9ERN3	Q9ERN3 rattus norv
39	171	6.5	95	11	Q9EP99	Q9EP99 rattus norv
40	169	6.5	372	6	Q28285	Q28285 canis fami
41	167	6.4	2310	16	Q8CMU7	Q8CMU7 staphylococ
42	165	6.3	786	5	Q21027	Q21027 caenorhabdi
43	164.5	6.3	631	11	Q93K60	Q93K60 mus musculu
44	162.5	6.2	2772	5	Q9VAV4	Q9VAV4 drosophila
45	161	6.2	1275	5	Q76602	Q76602 caenorhabdi

## ALIGNMENTS

RESULT 1  
ID Q86227 PRELIMINARY; PRT; 493 AA.  
AC Q86227;  
CD 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CD44 antigen.  
GN CD44.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP TISSUE=Colon adenocarcinoma;  
RC Wiese G.J., Freund D., Corbell D.,  
RA "Sequence analysis of the human CD44 antigen."  
RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AY101192; AF550040.1; -  
DR GO:GO:0016020; C:membrane; IEA.  
DR GO:GO:0004895; F:cell adhesion receptor activity; IEA.  
DR GO:GO:0005540; F:hyaluronic acid binding; IEA.  
DR GO:GO:0007155; F:cell adhesion; IEA.  
DR InterPro: IPR001231; CD44 antigen.  
DR InterPro: IPR000538; Link.  
DR Pfam: PF00193; Link; 1.  
DR PRINTS: PR00658; CD44.  
DR PRINTS: PR01265; LINKMODULF.  
DR PRODOM: PD00918; Link; 1.  
DR SMART: SM00445; Link; 1.  
DR PROSITE: PS01241; Link; 1.  
SQ SEQUENCE 493 AA; 53398 MW; 9C59CCCC8C261E6 CRC64;

Query Match Similarity 100.0%; Score 2608; DB 4; Length 493;  
Best Local Similarity 100.0%; Pred. No. 3.9e-177; Indels 0; Gaps 0;  
Matches 493; Conservative 0; Mismatches 0;

QY 1 MDKFWHAAWGLCLVPLSLAQLDILITCRFAGVFHVERNGRYSISRTAADLCAFNSTL 60

DB 1 MDKFWMAAMGLCLVPLSLAQIDINTCRFAGVFEKNGRYSISREADLCKAFNSTL 60  
QY 61 PTMAOMKALISIGETCRGFIIEGHVVIPIIHNSICANNIGVYIITNSTQDYTCFN 120  
DB 61 PTMAOMKALISIGETCRGFIIEGHVVIPIIHNSICANNIGVYIITNSTQDYTCFN 120  
QY 121 ASAPPEEDCTSVTLPAFPGPITITIVNDGTRVYOKGXYRINPEDIYPSNPTDDVSS 180  
DB 121 ASAPPEEDCTSVTLPAFPGPITITIVNDGTRVYOKGXYRINPEDIYPSNPTDDVSS 180  
QY 181 GSSSERSTSGGYIYFTFSTVHPIDPDSFWITDSTRIPATNMDSSHSTTLQPTANPT 240  
DB 181 GSSSERSTSGGYIYFTFSTVHPIDPDSFWITDSTRIPATNMDSSHSTTLQPTANPT 240  
QY 241 GLVEDLDRGTPLSMITQOQNSQSFSTSHGLEEDKOPHTSTLSSNRADYVGRDPMH 300  
DB 241 GLVEDLDRGTPLSMITQOQNSQSFSTSHGLEEDKOPHTSTLSSNRADYVGRDPMH 300  
QY 301 SEGSTLLEGYTSHPPTKESRTFIPVTSKATGSPGVAVTVGDSNSVNRSLSGDQTF 360  
DB 301 SEGSTLLEGYTSHPPTKESRTFIPVTSKATGSPGVAVTVGDSNSVNRSLSGDQTF 360  
QY 361 HPSGSGHTTGSEDSHSGSQEGGANTTSGPIRTPOIPFWLIIASLALAILAVCIA 420  
DB 361 HPSGSGHTTGSEDSHSGSQEGGANTTSGPIRTPOIPFWLIIASLALAILAVCIA 420  
QY 421 VNSRRRCGQKKQVLVINGNGAVEDRKPSGLNGEASKSQEMVHLVKNESSTPDQMTADE 480  
DB 421 VNSRRRCGQKKQVLVINGNGAVEDRKPSGLNGEASKSQEMVHLVKNESSTPDQMTADE 480  
QY 481 TRNLQNVDMKIGV 493  
DB 481 TRNLQNVDMKIGV 493

## RESULT 2

Q96J24 PRELIMINARY; PRT; 699 AA.  
AC Q96J24  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Similar to CD44 antigen (Homing function and Indian blood group system).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
EMBL; BC004372; AA04372.1; -  
DR PIR; I37369; I37369.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:004895; F:cell adhesion receptor activity; IEA.  
DR GO; GO:000540; F:hyaluronic acid binding; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR001231; CD44\_antigen.  
DR InterPro; IPR000538; Link.  
DR Pfam; PF00193; Xlink; 1.  
DR PRINTS; PR00658; CD44.  
DR PRODOM; PD000918; Link; 1.  
DR SMART; SM00445; Link; 1.  
DR SMART; PS01241; Link; 1.  
SQ SEQUENCE 699 AA; 76628 MW; 68AC36A932DBC7C CMC64;

Query Match 95.5%; Score 2490; DB 4; Length 699;  
Best Local Similarity 70.4%; Pred. No. 1.5e-168;  
Matches 492; Conservative 1; Mismatches 0; Indels 206; Gaps 1;

QY 1 MDKFWMAAMGLCLVPLSLAQIDINTCRFAGVFEKNGRYSISREADLCKAFNSTL 60  
DB 1 MDKFWMAAMGLCLVPLSLAQIDINTCRFAGVFEKNGRYSISREADLCKAFNSTL 60  
QY 61 PTMAOMKALISIGETCRGFIIEGHVVIPIIHNSICANNIGVYIITNSTQDYTCFN 120  
DB 61 PTMAOMKALISIGETCRGFIIEGHVVIPIIHNSICANNIGVYIITNSTQDYTCFN 120  
QY 121 ASAPPEEDCTSVTLPAFPGPITITIVNDGTRVYOKGXYRINPEDIYPSNPTDDVSS 180  
DB 121 ASAPPEEDCTSVTLPAFPGPITITIVNDGTRVYOKGXYRINPEDIYPSNPTDDVSS 180  
QY 181 GSSSERSTSGGYIYFTFSTVHPIDPDSFWITDSTRIPATNMDSSHSTTLQPTANPT 240  
DB 181 GSSSERSTSGGYIYFTFSTVHPIDPDSFWITDSTRIPATNMDSSHSTTLQPTANPT 240  
QY 241 GLVEDLDRGTPLSMITQOQNSQSFSTSHGLEEDKOPHTSTLSSNRADYVGRDPMH 300  
DB 241 GLVEDLDRGTPLSMITQOQNSQSFSTSHGLEEDKOPHTSTLSSNRADYVGRDPMH 300  
QY 301 SEGSTLLEGYTSHPPTKESRTFIPVTSKATGSPGVAVTVGDSNSVNRSLSGDQTF 360  
DB 301 SEGSTLLEGYTSHPPTKESRTFIPVTSKATGSPGVAVTVGDSNSVNRSLSGDQTF 360  
QY 361 HPSGSGHTTGSEDSHSGSQEGGANTTSGPIRTPOIPFWLIIASLALAILAVCIA 420  
DB 361 HPSGSGHTTGSEDSHSGSQEGGANTTSGPIRTPOIPFWLIIASLALAILAVCIA 420  
QY 421 VNSRRRCGQKKQVLVINGNGAVEDRKPSGLNGEASKSQEMVHLVKNESSTPDQMTADE 480  
DB 421 VNSRRRCGQKKQVLVINGNGAVEDRKPSGLNGEASKSQEMVHLVKNESSTPDQMTADE 480  
QY 481 TRNLQNVDMKIGV 493  
DB 481 TRNLQNVDMKIGV 493

## RESULT 3

Q9H5A6 PRELIMINARY; PRT; 470 AA.  
AC Q9H5A6  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE D68D18.2.5 (CD44 antigen (Homing function and Indian blood group system)) (Fragment).  
GN CD44.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Cobley V.;  
RA Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
EMBL; AL133330; CAC10346.1; -  
DR HSSP; P98066; ITSG.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:004895; F:cell adhesion receptor activity; IEA.  
DR GO; GO:000540; F:hyaluronic acid binding; IEA.

DR GO:0007155; P:cell adhesion; IEA.  
 DR InterPro; IPR001231; CD44 antigen.  
 DR InterPro; IPR000538; Link.  
 DR Pfam; PF00193; Xlink; 1.  
 DR PRINTS; PR00658; CD44.  
 DR PRINTS; PR01265; LINKMODULE.  
 DR PRODOM; PD000918; Link; 1.  
 DR SMART; SM00445; Link; 1.  
 DR SMART; SM00445; Link; 1.  
 DR PROSITE; PS01241; Link; 1.  
 FT NON TER 1  
 SQ SEQUENCE 470 AA; 50727 MW; E4B47CF12500D568 CRC64;  
  
 Query Match 94.6%; Score 2466; DB 4; Length 470;  
 Best Local Similarity 99.8%; Pred. No. 4.5e-167;  
 Matches 469; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 24 LITICRFAGVFEHKNKGRYSISRTTEADLCKAFNSTLPTMAQMEKALSIGFETCRYGFI 83  
 DB 1 LITICRFAGVFEHKNKGRYSISRTTEADLCKAFNSTLPTMAQMEKALSIGFETCRYGFI 60  
 QY 84 GHVVIPIRHPNSICANNNGVYILTSNTSQDYTYCFNAPPEEDCTSVTDLPNAPDGI 143  
 DB 61 GHVVIPIRHPNSICANNNGVYILTSNTSQDYTYCFNAPPEEDCTSVTDLPNAPDGI 120  
 QY 144 TITLVNRDGRVYQKGRFRTNPEDIPSNPTDDVSSGSSSSSTSGGYIFPTSTVHP 203  
 DB 121 TITLVNRDGRVYQKGRFRTNPEDIPSNPTDDVSSGSSSSSTSGGYIFPTSTVHP 180  
 QY 204 IPEDSPWITDSTRIPATNMDSSHSTLLOPTANPTGLVEDLDRTGPLSMTOQSNQS 263  
 DB 181 IPEDSPWITDSTRIPATNMDSSHSTLLOPTANPTGLVEDLDRTGPLSMTOQSNQS 240  
 QY 264 FSTSHBELDKDHPHTSTLTSSNRNDVYGGRRDPNHSSESTLLAGTSHYPTKESRT 323  
 DB 241 FSTSHBELDKDHPHTSTLTSSNRNDVYGGRRDPNHSSESTLLAGTSHYPTKESRT 300  
 QY 324 FIPVTSKTSFGVAVTVGDSNSNVNRSLSGDDPTFHPGSGHTTHGSEDSHSGOE 383  
 DB 301 FIPVTSKTSFGVAVTVGDSNSNVNRSLSGDDPTFHPGSGHTTHGSEDSHSGOE 360  
 QY 384 GQANTSGPIRTPOIPFWLIIILASLALAILAVCIANVSRRCGGKKLVINSNGAVE 443  
 DB 361 GQANTSGPIRTPOIPFWLIIILASLALAILAVCIANVSRRCGGKKLVINSNGAVE 420  
 QY 444 DRKPSGLNGEASKSQEMVHLVKNKSESTPDQFMTADETRLQNVDMKIGV 493  
 DB 421 DRKPSGLNGEASKSQEMVHLVKNKSESTPDQFMTADETRLQNVDMKIGV 470  
  
 RESULT 4  
 Q9UJ36 PRELIMINARY; PRT; 742 AA.  
 ID Q9UJ36  
 AC Q9UJ36  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transmembrane glycoprotein precursor.  
 GN CD44.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gunther U.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94147793; Pubmed=7508842;  
 RA Gunther U.;  
 RT "CD44: a multitude of isoforms with diverse functions";  
 RL Curr. Top. Microbiol. Immunol. 184:47-63 (1993).  
 DR EMBL; AJ251595; CAB61878.1; -.

DR PIR; A47195; A47195.  
 DR HSSP; P98066; ITSG.  
 DR GO:0016020; C:membrane.  
 DR GO:0004885; F:cell adhesion receptor activity; IEA.  
 DR GO:0005440; F:hyaluronic acid binding; IEA.  
 DR GO:0007155; P:cell adhesion; IEA.  
 DR InterPro; IPR001231; CD44 antigen.  
 DR InterPro; IPR000538; Link.  
 DR Pfam; PF00193; Xlink; 1.  
 DR PRINTS; PR00658; CD44.  
 DR PRODOM; PD000918; Link; 1.  
 DR SMART; SM00445; Link; 1.  
 DR PROSITE; PS01241; Link; 1.  
 KW Signal.  
 FT CHAIN 1 23  
 SQ SEQUENCE 742 AA; 81598 MW; C319E5CE50B51D3 CRC64;  
  
 Query Match 94.2%; Score 2457.5; DB 4; Length 742;  
 Best Local Similarity 66.0%; Pred. No. 3.2e-166;  
 Matches 490; Conservative 2; Mismatches 1; Indels 249; Gaps 1;  
  
 QY 1 MDKFWMAAGCLVPLSLAQIDNTICRFAGVFEHKNKGRYSISRTTEADLCKAFNSTL 60  
 DB 1 MDKFWMAAGCLVPLSLAQIDNTICRFAGVFEHKNKGRYSISRTTEADLCKAFNSTL 60  
 QY 61 PTMAQMEKALSIGFETCRYGFI 120  
 DB 61 PTMAQMEKALSIGFETCRYGFI 120  
 QY 121 ASAPPEEDCTSVTDLPNAPDGI 180  
 DB 121 ASAPPEEDCTSVTDLPNAPDGI 180  
 QY 181 GSSSERSTSGGYIFPTSTVHP 222  
 DB 181 GSSSERSTSGGYIFPTSTVHP 222  
 QY 241 AMDWFSWFLPSESKNHLHTTQMAGTSSNTISAGWEPNEBDEDRHLSPSSGGIDDD 300  
 DB 241 AMDWFSWFLPSESKNHLHTTQMAGTSSNTISAGWEPNEBDEDRHLSPSSGGIDDD 300  
 QY 223  
 DB 223  
 QY 301 EDPISTISTTPRAFDHTKQNDWTQNNPNSNEVLLQTTMTVDVANGTAYEGNMN 360  
 DB 301 EDPISTISTTPRAFDHTKQNDWTQNNPNSNEVLLQTTMTVDVANGTAYEGNMN 360  
 QY 223  
 DB 223  
 QY 361 PEAPPLIHHEHBEETPHSTSTIQATPSSTEETATOKOWFGNWHGYNQTPREDS 420  
 DB 223  
 QY 421 HSTTGTAASAHSHHPQGRTPPEEDSWTDFNPIISHMGRGHQGRMDMDSSSTT 480  
 DB 421 HSTTGTAASAHSHHPQGRTPPEEDSWTDFNPIISHMGRGHQGRMDMDSSSTT 480  
 QY 232 LQPTANPTGLVEDLDRTGPLSMTOQSNQSGSTHGELEDKHPHTSTLTSSNRNDV 291  
 DB 232 LQPTANPTGLVEDLDRTGPLSMTOQSNQSGSTHGELEDKHPHTSTLTSSNRNDV 291  
 QY 481 LQPTANPTGLVEDLDRTGPLSMTOQSNQSGSTHGELEDKHPHTSTLTSSNRNDV 540  
 DB 481 LQPTANPTGLVEDLDRTGPLSMTOQSNQSGSTHGELEDKHPHTSTLTSSNRNDV 540  
 QY 292 TGGRRDPNHSSESTLLAGTSHYPTKESRTFIPVTSKTSFGVAVTVGDSNSNVN 351  
 DB 292 TGGRRDPNHSSESTLLAGTSHYPTKESRTFIPVTSKTSFGVAVTVGDSNSNVN 351  
 QY 541 TGGRRDPNHSSESTLLAGTSHYPTKESRTFIPVTSKTSFGVAVTVGDSNSNVN 600  
 DB 541 TGGRRDPNHSSESTLLAGTSHYPTKESRTFIPVTSKTSFGVAVTVGDSNSNVN 600  
 QY 352 SLSDGDDPTFHPGSGHTTHGSEDSHSGOEAGANTSGPIRTPOIPFWLIIILASLAL 411  
 DB 352 SLSDGDDPTFHPGSGHTTHGSEDSHSGOEAGANTSGPIRTPOIPFWLIIILASLAL 411  
 QY 601 SLSDGDDPTFHPGSGHTTHGSEDSHSGOEAGANTSGPIRTPOIPFWLIIILASLAL 660  
 DB 601 SLSDGDDPTFHPGSGHTTHGSEDSHSGOEAGANTSGPIRTPOIPFWLIIILASLAL 660  
 QY 412 ALIIAVCIANVSRRCGGKKLVINSNGAVEDEKPSGLNGEASKSQEMVHLVKNKSEST 471  
 DB 412 ALIIAVCIANVSRRCGGKKLVINSNGAVEDEKPSGLNGEASKSQEMVHLVKNKSEST 471  
 QY 661 ALIIAVCIANVSRRCGGKKLVINSNGAVEDEKPSGLNGEASKSQEMVHLVKNKSEST 720  
 DB 661 ALIIAVCIANVSRRCGGKKLVINSNGAVEDEKPSGLNGEASKSQEMVHLVKNKSEST 720  
 QY 472 PDQFMTADETRLQNVDMKIGV 493  
 DB 472 PDQFMTADETRLQNVDMKIGV 493

Db 721 PDQMTADETRNLQNVDMKIGV 742

## RESULT 5

Q9H5A7 PRELIMINARY; PRT; 676 AA.

AC Q9H5A7;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
DE D06BD18.2.4 (CD4 antigen (Homing function and Indian blood group system)) (Fragment).  
GN CD44.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA COBLEY V.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL133330; CAC10345.1; -.  
DR HSSP; P98066; ITSG.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.  
DR GO; GO:0005540; F:hyaluronic acid binding; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR001231; CD44 antigen.  
DR Pfam; PF00193; Xlink; 1.  
DR PRINTS; PR00658; CD44.  
DR PRODOM; PD000918; Link; 1.  
DR SMART; SM00445; Link; 1.  
DR PROSITE; PS01241; Link; 1.  
FT NON TER 1 1  
SQ SEQUENCE 676 AA; 73929 MW; 122C7250B7DA0F47 CRC64;

Query Match 90.0%; Score 2348; DB 4; Length 676;  
Best Local Similarity 69.2%; Pred. No. 1.7e-158; Indels 206; Gaps 1;  
Matches 468; Conservative 1; Mismatches 1;

QY 24 LNTICRFAGVHVEKNGRYSISRTBADLCAFNSTLPYVAQMEKALISIGETCRYGIE 83  
Db 1 LNTICRFAGVHVEKNGRYSISRTBADLCAFNSTLPYVAQMEKALISIGETCRYGIE 60  
QY 84 GHVVIPIRHNSICANNNGVYILTSNTSOYDYTCFNASAPREEDCTSVTDLPNAPGPI 143  
Db 61 GHVVIPIRHNSICANNNGVYILTSNTSOYDYTCFNASAPREEDCTSVTDLPNAPGPI 120  
QY 144 TITIVNRDGRTRYVQGEYRTNPEDIYPSNPTDDVSSGSSSSRSSTGGYIFTFSTVHP 203  
Db 121 TITIVNRDGRTRYVQGEYRTNPEDIYPSNPTDDVSSGSSSSRSSTGGYIFTFSTVHP 180  
QY 204 IPDEDSPWITDSTRIPAT----- 222  
Db 181 IPDEDSPWITDSTRIPATSTSSNTISAGWEPNEENEDERHLFSGSGIDDEDFISS 240  
QY 223 ----- 222  
Db 241 TISTTPRAFDHTKQNDQWQNPSPNSPEVLLQTTMTVDVDRNGTAYEGNNPEAHNP 300  
QY 223 ----- 222  
Db 301 LIHHEHHEEETPHSTIGATPSSTEEATQKEQWGNRMHGYRQTPKEDSHSTTGT 360  
QY 223 ----- 222  
Db 361 AAASAHSHPMQGRTPPSPEDSWTFPNPISHMGHQAQGRMDSDSHSTIGTAN 420  
QY 238 PNTGLVEDLRTGSLSMTTQOSNSQSTSHSGLEEDKDPFTSTLTSSKRDVYTGGRD 237  
Db 421 PNTGLVEDLRTGSLSMTTQOSNSQSTSHSGLEEDKDPFTSTLTSSKRDVYTGGRD 480

QY 298 PNHSGSTTLLEGYTSHTYPTKESRTFIPYTSATGSGFYATAVGSDNSNVRSLSGDQ 357  
Db 481 PNHSGSTTLLEGYTSHTYPTKESRTFIPYTSATGSGFYATAVGSDNSNVRSLSGDQ 540  
QY 358 DTFPSCGSHHTHSGSDGSHSGQEGGANTTSGPIRTPOIPEMLIILASLALALITLAV 417  
Db 541 DTFPSCGSHHTHSGSDGSHSGQEGGANTTSGPIRTPOIPEMLIILASLALALITLAV 600  
QY 418 CIANRRRCGQCKKLVIYNSGNGAVEDRKPSGLNGEASKSOEYVHLVKNKSSSTPDQFMT 477  
Db 601 CIANRRRCGQCKKLVIYNSGNGAVEDRKPSGLNGEASKSOEYVHLVKNKSSSTPDQFMT 660  
QY 478 ADETRNLQNVDMKIGV 493  
Db 661 ADETRNLQNVDMKIGV 676

## RESULT 6

Q9H5A5 PRELIMINARY; PRT; 719 AA.

AC Q9H5A5;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
DE D06BD18.2.3 (CD4 antigen (Homing function and Indian blood group system)) (Fragment).  
GN CD44.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA COBLEY V.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL133330; CAC10347.1; -.  
DR HSSP; P98066; ITSG.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.  
DR GO; GO:0005540; F:hyaluronic acid binding; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR001231; CD44 antigen.  
DR Pfam; PF00193; Xlink; 1.  
DR PRINTS; PR00658; CD44.  
DR PRODOM; PD000918; Link; 1.  
DR SMART; SM00445; Link; 1.  
DR PROSITE; PS01241; Link; 1.  
FT NON TER 1 1  
SQ SEQUENCE 719 AA; 78854 MW; 36B747A896AD7380 CRC64;

Query Match 89.2%; Score 2326.5; DB 4; Length 719;  
Best Local Similarity 65.1%; Pred. No. 6.3e-157; Indels 249; Gaps 1;  
Matches 468; Conservative 1; Mismatches 1;

QY 24 LNTICRFAGVHVEKNGRYSISRTBADLCAFNSTLPYVAQMEKALISIGETCRYGIE 83  
Db 1 LNTICRFAGVHVEKNGRYSISRTBADLCAFNSTLPYVAQMEKALISIGETCRYGIE 60  
QY 84 GHVVIPIRHNSICANNNGVYILTSNTSOYDYTCFNASAPREEDCTSVTDLPNAPGPI 143  
Db 61 GHVVIPIRHNSICANNNGVYILTSNTSOYDYTCFNASAPREEDCTSVTDLPNAPGPI 120  
QY 144 TITIVNRDGRTRYVQGEYRTNPEDIYPSNPTDDVSSGSSSSRSSTGGYIFTFSTVHP 203  
Db 121 TITIVNRDGRTRYVQGEYRTNPEDIYPSNPTDDVSSGSSSSRSSTGGYIFTFSTVHP 180  
QY 204 IPDEDSPWITDSTRIPAT----- 222  
Db 181 IPDEDSPWITDSTRIPATSTSSNTISATRTATKQETWQFSLFLPSSKXHLHTTQ 240

Qy	223	-----	222
Db	241	MAGTSNNTISAGMEPHEENEDERDRILSSGSGIDDDDDFISSITSTTPRAFDHKKNO	3000
Qy	223	-----	222
Db	301	WTQWNPESHNPVELLQTTTRMTDVDNRNGTAYEGWNPENAPPLIHHEHEEBETPHSTS	3600
Qy	223	-----	222
Db	361	TIQATBSSTTEETATQKEOMFGKRMHEGRTQPKSDSSTTGTAALASHTHPMQGTTT	4200
Qy	223	-----	222
Db	421	SPEDSWTDFFNFI SHPMGRGHQAGRMDMSHSITLQTPANPTGLVEDLDTGPLSM	4800
Qy	223	-----	222
Db	481	TIQOOSNQSTSHHEGLEEDKOPHTSTLTSNRNDVYGRRDPHSHSGSTTLLEGYTS	5400
Qy	223	-----	222
Db	541	YPHTKESRTPIPTTSAKTGSFGYTAIVYDSSNNVRSLSGODTFRHSQGSHTTHSES	6000
Qy	223	-----	222
Db	601	DGSHSHSGOEGGANTTGGPIRTPOIPEMLIIIASLIALALIIAVCIAVNSRRRCQKKLV	6600
Qy	223	-----	222
Db	661	INGANGAVERKXSGLNGEASKSQEWVHVNSESETPDOFMTADETRNLONVMYKIGV	7200

```

RESULT 7
062913
ID AC PRELIMINARY; PRT; 514 AA.
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DS CD44i (Fragment) .
OS Rattus norvegicus (Rat) .
OC Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Chondrosarcoma;
RA Stevens J.W., Midura R.U.;
RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; U46958; AAA92921.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO: GO:0005540; F:hyaluronic acid binding; IEA.
DR GO: GO:0007155; P:cell adhesion; IEA.
DR InterPro: IPR001231; CD44_antigen.
DR InterPro: IPR000538; Link.
DR Pfam: PF00193; Xlink; 1.
DR PRINTS: PR00658; CD44.
DR PRINTS: PR01265; LINKMODULE.
DR Prodom: PD000918; Link; 1.
DR SMART: SM00445; Link; 1.
FT NON_TER 1
SQ SEQUENCE 514 AA; 55647 MW; 3744984FOE3DDA92 CRC64;
Query Match 84.0%; Score 2189.5; DB 11; Length 514;
Best Local Similarity 82.0%; Pred. No. 2.2e-147;
Matches 426; Conservative 1; Mismatches 0; Indels 87; Gaps 1
Cy 67 EKALSTGFEFCRCGFLIEGHVLPRIHPNSICAAANTGYVLTNSTSODVYCFAPASPE 126
Db 1 EKALSTGFEFCRCGFLIEGHVLPRIHPNSICAAANTGYVLTNSTSODVYCFAPASPE 60

```

Qy	127	EDCTSYTDLPNPNFDDGITITITIVNRGKTYVQKGEYRNPEDIPYSPNTPDDVSSGSSSER	186
Dd	61	EDCTSYTDLPNPNFDDGITITITIVNRGKTYVQKGEYRNPEDIPYSPNTPDDVSSGSSSER	120
Qy	187	SSTSGYIFYFTSYVHPIDPDEDSPIWIDSTDIRPAT-----	222
Dd	121	SSTSGYIFYFTSYVHPIDPDEDSPIWIDSTDIRPATIQATPSYTBETAQKQCFQFQNR	180
Qy	223	-----	222
Dd	181	HGCIROTPEBDHSHTTGAAASATSHMQORITTPSPEDBSWIDFNPJISHPMGRGHQAG	240
Qy	223	---NDSDSHSTLQETANPNPTGLVEDLDRTPLSMTTQOSNSQSFSTSHGLEEDXDHPT	279
Dd	241	RRMDMSHSTTLQETANPNPTGLVEDLDRTPLSMTTQOSNSQSFSTSHGLEEDXDHPT	300
Qy	280	TSTLTSSNNVDYGRRDPNHSBGSTLLBSGYTHYPHTKESRFIVTSAKTGSGYTA	339
Dd	301	TSTLTSSNNVDYGRRDPNHSBGSTLLBSGYTHYPHTKESRFIVTSAKTGSGYTA	360
Qy	340	VYVGDNSNVNRSLSGDDOTFHPSCGSHTTGSESDGSHSQCBGANTTSGPIRTPOLP	399
Dd	361	VYVGDNSNVNRSLSGDDOTFHPSCGSHTTGSESDGSHSQCBGANTTSGPIRTPOLP	420
Qy	400	EWIITIASLALALILAVCIAVNSRRCGQKKXIVINSNGAVEDRXPGLNGEBAKXOE	459
Dd	421	EWIITIASLALALILAVCIAVNSRRCGQKKXIVINSNGAVEDRXPGLNGEBAKXOE	480
Qy	460	MVHLVNXKSESPTPDQPMTADEFBNLOANDMKIGV	493
Dd	481	MVHLVNXKSESPTPDQPMTADEFBNLOANDMKIGV	514

RESULT 8	0867T2	PRELIMINARY;	PRT;	361 AA.
ID	0867T2			
AC	0867T2;			
DT	01-JUN-2003 (T-EMBLrel. 24, Created)			
DT	01-JUN-2003 (T-EMBLrel. 24, Last sequence update)			
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)			
DE	Hypothetical protein DKFZp451K1918.			
GN	DKFZp451K1918.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,			
RA	Osanger A., Wiemann S.;			
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AL832642; CDS8965.1; -			
DR	GO: GO:0016020; Cmembrane; IEA.			
DR	GO: GO:0004895; F-cell adhesion receptor activity; IEA.			
DR	GO: GO:0005540; P-hyaluronic acid binding; IEA.			
DR	GO: GO:0007155; P-cell adhesion; IEA.			
DR	InterPro: IPR001231; CD44 antigen.			
DR	InterPro: IPR000538; Link.			
DR	Pfam: PF00193; Link; 1.			
DR	PRINTS: PRO0658; Link.			
DR	PRINTS: PRO1265; LINKMODULE.			
DR	PRODom: PD000918; Link; 1.			
DR	SMART: SM00445; Link; 1.			
DR	PROSITE: PS01241; Link; 1.			
KW	Hypothetical protein.			
SQ	SEQUENCE 361 AA; 39387 MW; F51A746B4387C54 CRC64;			
Query March	70.4%;	Score 1837;	DB 4;	Length 361;
Best Local Similarity	72.8%;	Pred. No. 1,6e-122;		
Matches 35;	Conservative 0;	Mismatches 2;	Indels 132;	Gaps 1

```
Db 1 MDKFWHAAAGLCLVPLSLAQIDLNTCRFAGVHFKNGRYSISRTAADLCKAFNSTL 60
Qy 61 PTMAOMEKALSIGETCRGYFIEGHVYIPRIHPNSICAAANTGVYILTSNTSQDYTCFN 120
Db 61 PTMAOMEKALSIGETCRGYFIEGHVYIPRIHPNSICAAANTGVYILTSNTSQDYTCFN 120
Qy 121 ASAPPEEDCTSVTDLPNAPFGPITITIVNRDGRYVQKGEYRNPEDIPSNPTDDVSS 180
Db 121 ASAPPEEDCTSVTDLPNAPFGPITITIVNRDGRYVQKGEYRNPEDIPSNPTDDVSS 180
Qy 181 GSSSERSTSGGYFYFTFSYVHPPIPDSDSPMTDSTRIPATMDSHSTTLQPTANPNT 240
Db 181 GSSSERSTSGGYFYFTFSYVHPPIPDSDSPMTDSTRIPATMDSHSTTLQPTANPNT 240
Qy 241 GLVEDLDRGTGPLSMTTQOSNSQSFSTSHGLEDKCHPTTSTLTSSNRNDVTGGRDPNH 300
Db 241 GLVEDLDRGTGPLSMTTQOSNSQSFSTSHGLEDKCHPTTSTLTSSNRNDVTGGRDPNH 300
Qy 301 SEGSTLLEGYTSHYPTKSRPIVYSAKTSFGYAVTVGDSNVNRSLSGDQDTF 360
Db 301 SEGSTLLEGYTSHYPTKSRPIVYSAKTSFGYAVTVGDSNVNRSLSGDQDTF 360
Qy 361 HPSGSHHTHGESDGHSHGSEGGANTSGPIRTPOIPFWLIIASLALALILAVCIA 420
Db 229 HPSGSHHTHGESDGHSHGSEGGANTSGPIRTPOIPFWLIIASLALALILAVCIA 288
Qy 421 VNSRRRCGQKKLVYNSGNAVEDRKPSGLNGEASKSQEVHLVKNESSETPQFMFADE 480
Db 289 VNSRRRCGQKKLVYNSGNAVEDRKPSGLNGEASKSQEVHLVKNESSETPQFMFADE 348
Qy 481 TRNLQNVDMKIGV 493
Db 349 TRNLQNVDMKIGV 361
```

## RESULT 9

Q8N694

PRELIMINARY; PRT; 361 AA.

AC Q8N694;

DT 01-OCT-2002 (Tremblrel. 22, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DN CD44 antigen.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OC NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Normal pigmented retinal epithelium;

RT "Sequence analysis of the human CD44 antigen.";

RT Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.

DR EMBL; AY101193; AAM5041.1; -

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005540; F:hyaluronic acid binding; IEA.

DR GO; GO:007155; F:cell adhesion; IEA.

DR InterPro; IPR001231; CD44\_antigen.

DR Pfam; PF00193; Xlink; 1.

DR PRINTS; PRO0658; CD44.

DR PRINTS; PRO1265; LINKMODULE.

DR Prodom; PD000918; Link; 1.

DR SMART; SM00445; LINK; 1.

DR PROSITE; PS01241; LINK; 1.

SQ SEQUENCE 361 AA; 39442 MW; 597A746B43331C0 CRC64;

```
Qy 1 MDKFWHAAAGLCLVPLSLAQIDLNTCRFAGVHFKNGRYSISRTAADLCKAFNSTL 60
Db 1 MDKFWHAAAGLCLVPLSLAQIDLNTCRFAGVHFKNGRYSISRTAADLCKAFNSTL 60
Qy 61 PTMAOMEKALSIGETCRGYFIEGHVYIPRIHPNSICAAANTGVYILTSNTSQDYTCFN 120
Db 61 PTMAOMEKALSIGETCRGYFIEGHVYIPRIHPNSICAAANTGVYILTSNTSQDYTCFN 120
Qy 121 ASAPPEEDCTSVTDLPNAPFGPITITIVNRDGRYVQKGEYRNPEDIPSNPTDDVSS 180
Db 121 ASAPPEEDCTSVTDLPNAPFGPITITIVNRDGRYVQKGEYRNPEDIPSNPTDDVSS 180
Qy 181 GSSSERSTSGGYFYFTFSYVHPPIPDSDSPMTDSTRIPATMDSHSTTLQPTANPNT 240
Db 181 GSSSERSTSGGYFYFTFSYVHPPIPDSDSPMTDSTRIPATMDSHSTTLQPTANPNT 240
Qy 241 GLVEDLDRGTGPLSMTTQOSNSQSFSTSHGLEDKCHPTTSTLTSSNRNDVTGGRDPNH 300
Db 241 GLVEDLDRGTGPLSMTTQOSNSQSFSTSHGLEDKCHPTTSTLTSSNRNDVTGGRDPNH 300
Qy 301 SEGSTLLEGYTSHYPTKSRPIVYSAKTSFGYAVTVGDSNVNRSLSGDQDTF 360
Db 301 SEGSTLLEGYTSHYPTKSRPIVYSAKTSFGYAVTVGDSNVNRSLSGDQDTF 360
Qy 361 HPSGSHHTHGESDGHSHGSEGGANTSGPIRTPOIPFWLIIASLALALILAVCIA 420
Db 229 HPSGSHHTHGESDGHSHGSEGGANTSGPIRTPOIPFWLIIASLALALILAVCIA 288
Qy 421 VNSRRRCGQKKLVYNSGNAVEDRKPSGLNGEASKSQEVHLVKNESSETPQFMFADE 480
Db 289 VNSRRRCGQKKLVYNSGNAVEDRKPSGLNGEASKSQEVHLVKNESSETPQFMFADE 348
Qy 481 TRNLQNVDMKIGV 493
Db 349 TRNLQNVDMKIGV 361
```

## RESULT 10

Q8OX37

PRELIMINARY; PRT; 580 AA.

AC Q8OX37;

DT 01-JUN-2003 (Tremblrel. 24, Created)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DN Hypothetical protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=FVB/N; TISSUE=Salivary gland;

RT MEDLINE=22386257; PubMed=12477932;

RA Klausner R.D., Collins F.S., Wagner L., Shamen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.J.,

RA Rana S.S., Loquellano N.A., Peters G.J., Arramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E.,

RA Jones S.J., Meira W.A.,

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).



RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=salivary gland;  
 RA Strauberg R.;  
 RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC051388; AHS1388.1;  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR GO: GO:0004895; F:cell adhesion receptor activity; IEA.  
 DR GO: GO:0005540; F:hyaluronic acid binding; IEA.  
 DR GO: GO:0007155; P:cell adhesion; IEA.  
 DR InterPro: IPR001231; CD44 antigen.  
 DR InterPro: IPR00538; Link.  
 DR Pfam: PF00193; Link.  
 DR PRINTS: PR00658; CD44.  
 DR PRINTS: PR01265; LINKMODULE.  
 DR ProDom: PD000918; Link; 1.  
 DR SMART: SM00445; Link; 1.  
 DR PROSITE: PS01241; Link; 1.  
 KR Hypothetical protein  
 SQ SEQUENCE 360 AA; 63264 MW; F38A59C3AAB8BA0F CRC64;  
 Query Match 67.0%; Score 1748.5; DB 11; Length 580;  
 Best Local Similarity 63.2%; Pred. No. 5.6e-116;  
 Matches 368; Conservative 38; Mismatches 85; Indels 91; Gaps 10;  
 QY 1 MDKRWMAAGLCIVPLSLA-----QIDNITCRFAGVFRHKKGRYSISRTKADLCAP 56  
 DB 1 MDKRWMAAGLCIVPLSLA-----QIDNITCRFAGVFRHKKGRYSISRTKADLCAP 60  
 QY 57 NSTLPTAOMEKALISFETCRVYFIEGHVAVIRIHNSICANNNTGYIL-TSNTSOYD 115  
 DB 61 NSTLPTDOKMLAKSKFETCRVYFIEGHVAVIRIHNSICANNNTGYILVTSNTHD 120  
 QY 116 TYCNASAPREDCTSTVDLPNAPDGPITITVNRDCTRVYQGEHRTNEDYPSNPFD 175  
 DB 121 TYCNASAPREDCTSTVDLPNSPDGPITITVNRDCTRVYQGEHRTNEDYPSNPFD 180  
 QY 176 DVSSGSSERSSTSGGIFFYTF-----STVH----- 202  
 DB 181 DVSSGSSERSSTSGGIFFYTF-----STVH----- 202  
 QY 203 -----PIPEDS-----PWTTSSTRI----- 219  
 DB 240 WFGQWQGNKPTSESDSHVTEGTTASAHNNHPSQRTTOSQEDVSV-TDFPPIHPMG 298  
 QY 220 -----PATNDSSTLTQPTANPTGVLVEDLRTGFLSMTQOSNGSSTSGHELEED 274  
 DB 299 QGHOTESKDTDSSTLTQPTANPTGVLVEDLRTGFLSMTQOSNGSSTSGHELEED 358  
 QY 275 KDHPSTLTLSNNDVYTGRRDPNHSBGSTLLGGTSHYPHTKSRFTIPVTSATGS 334  
 DB 359 ENHPTSTLTSPSTSGKACDARRGSLPTDPTTSVEGYTFQYDPDMEGTLFPVPAKTEV 418  
 QY 335 FGVAIVYVG-DSNNVNRSLSGDDTFHSGGSH--TTHGSESDGSHSGOEGANTSG 391  
 DB 419 FGEIVLTATDSNVVDSLPGRDSDSGSRRTYTHSGELAGHSANODSVTTTSG 478  
 QY 392 PIRTPQIPDWLIIASLALALILAVCIANVSRRCQKKKLVINSNGAVEDEKPSGLN 451  
 DB 479 PMRRPQIPDWLIIASLALALILAVCIANVSRRCQKKKLVINSNGAVEDEKPSGLN 538  
 QY 452 GSAKSGEMVHLVNRKESSTPDQFMATDETRNLQNVMKIGV 493  
 DB 539 GSAKSGEMVHLVNRKESSTPDQFMATDETRNLQNVMKIGV 580

DE DUG8D18.2.2 (CD44 antigen (Homing function and Indian blood group  
 DE system)) (Fragment).  
 GN CD44.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cobley V.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL133330; CAC10348.1; -  
 DR HSSP: P98066; ITSG.  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR GO: GO:0004895; F:cell adhesion receptor activity; IEA.  
 DR GO: GO:0005540; F:hyaluronic acid binding; IEA.  
 DR GO: GO:0007155; P:cell adhesion; IEA.  
 DR InterPro: IPR001231; CD44 antigen.  
 DR InterPro: IPR00538; Link.  
 DR Pfam: PF00193; Link; 1.  
 DR PRINTS: PR00658; CD44.  
 DR PRINTS: PR01265; LINKMODULE.  
 DR ProDom: PD000918; Link; 1.  
 DR SMART: SM00445; Link; 1.  
 DR PROSITE: PS01241; Link; 1.  
 KR NON TER  
 SQ SEQUENCE 338 AA; 36732 MW; 9835C309B692E4DB CRC64;  
 Query Match 65.4%; Score 1705; DB 4; Length 338;  
 Best Local Similarity 71.7%; Pred. No. 3.5e-113;  
 Matches 337; Conservative 0; Mismatches 1; Indels 132; Gaps 1;  
 QY 24 INITCRFAGVFRHKKGRYSISRTKADLCAPNSTLPTMAOMEKALISFETCRVYFIE 83  
 DB 1 INITCRFAGVFRHKKGRYSISRTKADLCAPNSTLPTMAOMEKALISFETCRVYFIE 60  
 QY 84 GHVAVIRIHNSICANNNTGYILTSNTSOYDYTCFNASAPREDCTSTVDLPNAPDGP 143  
 DB 61 GHVAVIRIHNSICANNNTGYILTSNTSOYDYTCFNASAPREDCTSTVDLPNAPDGP 120  
 QY 144 TITVNRDCTRVYQGEHRTNEDYPSNPFDSDVSSGSSERSSTSGGIFFYFSTVHP 203  
 DB 121 TITVNRDCTRVYQGEHRTNEDYPSNPFDSDVSSGSSERSSTSGGIFFYFSTVHP 180  
 QY 204 IPEDSPWITDSTDRIPATNDSSTLTQPTANPTGVLVEDLRTGFLSMTQOSNGS 263  
 DB 181 IPEDSPWITDSTDRIPATNDSSTLTQPTANPTGVLVEDLRTGFLSMTQOSNGS 200  
 QY 264 FSTHSGLEEDKHHTTSLTSNNDVYTGRRDPNHSBGSTLLGGTSHYPHTKSRFT 323  
 DB 201 ----- 200  
 QY 324 FIPVTSATGSFGVAVTVGDSNVNRSLSGDDTFHSGGSHHTHSGSDGSHSGOE 383  
 DB 201 -----DQTFHPSGGSHHTHSGSDGSHSGOE 228  
 QY 384 GGANTTSGPRTQIPDWLIIASLALALILAVCIANVSRRCQKKKLVINSNGAVE 443  
 DB 229 GGANTTSGPRTQIPDWLIIASLALALILAVCIANVSRRCQKKKLVINSNGAVE 288  
 QY 444 DRKPSGLNBSKSGEMVHLVNRKESSTPDQFMATDETRNLQNVMKIGV 493  
 DB 289 DRKPSGLNBSKSGEMVHLVNRKESSTPDQFMATDETRNLQNVMKIGV 338

RESULT 11  
 ID 09H5A4 PRELIMINARY; PRT; 338 AA.  
 AC 09H5A4;  
 DT 01-VAR-2001 (TREMBlrel. 16, Created)  
 DT 01-VAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

RESULT 12  
 ID 008779 PRELIMINARY; PRT; 780 AA.  
 AC 008779;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE CD44 protein.

GN CD44.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BX; TISSUE=Pancreas;  
 RA Holmann M.;  
 RT "Rattus norvegicus CD44 protein sequence."  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U96138; AB54002.1; -.  
 DR HSSP; P98066; ITRG.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.  
 DR GO; GO:0005540; F:hyaluronic acid binding; IEA.  
 DR GO; GO:0007155; P:cell adhesion; IEA.  
 DR InterPro; IPR001231; CD44\_antigen.  
 DR InterPro; IPR000538; Link.  
 DR Pfam; PF00193; Xlink; 1.  
 DR PRINTS; PR00658; CD44.  
 DR PRINTS; PR01265; LINKMODULE.  
 DR ProDom; PD000918; Link; 1.  
 DR SMART; SM00445; Link; 1.  
 DR PROSITE; PS01241; Link; 1.  
 DR SEQUENCE 780 AA; 85917 MW; CC4D35AB1EA7377C CRC64;

Query Match 62.8%; Score 1637.5; DB 11; Length 780;  
 Best Local Similarity 47.2%; Pred. No. 6.3e-108; Indels 295; Gaps 12;  
 Matches 370; Conservative 35; Mismatches 84;

QY 1 MDKFWMAWG-LCLVPLSLA--QIDNITCRFAGVHVEKNRGYSISRTAADLCKAFN 57  
 DB 1 MDKFWMAWG-LCLVPLSLA--QIDNITCRFAGVHVEKNRGYSISRTAADLCKAFN 60  
 QY 58 STPLTMAQEKALISGFETCRGFIIEGHVIVPRHPNSICANNVTGYI-LTNTSGOYDT 116  
 DB 58 STPLTMAQEKALISGFETCRGFIIEGHVIVPRHPNSICANNVTGYI-LTNTSGOYDT 116  
 DB 61 TLLPTMAQEKALISGFETCRGFIIEGHVIVPRHPNSICANNVTGYI-LTNTSGOYDT 120  
 QY 117 YCFNAPAPBEDCTSVTDLPNAPFDGPIITITVNRDGRVYQKGEYRTNPEIDYPSNFTD 176  
 DB 121 YCFNAPAPBEDCTSVTDLPNAPFDGPIITITVNRDGRVYQKGEYRTNPEIDYPSNFTD 180  
 QY 177 DVSSGSSSRSSSTSGGYTFYF-----STVHPIDED 208  
 DB 177 DVSSGSSSRSSSTSGGYTFYF-----STVHPIDED 208  
 DB 181 DVSSGSSSTIK-STPEGYTLHTDLPSTQPTGDRDAFIIGTLATLASTVYSKHATAQKQ 239  
 QY 209 SPWL----- 212  
 DB 240 NNWIMSWFNGSGSTTQTQSDPTTATATMTPTPPKQEAQNWFSMFQPSKSHLH 299  
 QY 213 -----TDSITRIP----- 220  
 DB 300 TTTKMPGTESNTNPTGKPNENEDTDKYPNFGSGGIDDEDFISSLTATTPWVSAHTK 359  
 QY 221 -----ATNND----- 225  
 DB 360 QNERTQWNPISNPEVLQTTTNTDIDRNSTASHGENTWQEPQPPFNHBYQDEETP 419  
 QY 226 ----- 225  
 DB 420 HATSTTWADPNSTTEBAATQKEKWFENWQGNPPTPSEDSHVTEGTASAHNNHPSQRM 479  
 QY 226 -----SSHTTLLQPTANPTGLVEDLDRGTPL 252  
 DB 480 TTQSQEDVSWTDFDPDISHPMQGHQTESKDTGSHSTTLQPTANPTGLVEDLDRGTPL 539  
 QY 253 SMTTQGSNSQSFSTSHGIEEDKDPHTTSTLTSNRNDVYTGKRPDNPNSBSSTLLAGYT 312  
 DB 540 SVTTPQSHQNSSTLPGLEBEDHPHTTSVLPSSYK--SGRRRGGSLPDDTTSLSGYT 596  
 QY 313 SHYPTKEGRFTIPVTSKATGSGFVTAATVG--DSNSNVRSLSGDQ--TFAPSGASHT-T 369

DB 597 PQYEDIMENGTLPFVTPAKTEVFGETGTAVATDSNFVDSLPQDGGSSMDPRGGFDTVT 656  
 QY 370 HGSISDGHSHSGQEGGANTTSGPRTQIPEMTLLASLLALAILAVCIANVSRRCQ 429  
 DB 657 HGSISLHSGNSGNSGVTTTSGPARPQIPEMTLLASLLALAILAVCIANVSRRCQ 716  
 QY 430 KKKLVINGANGAVDRKPSGNGEASKSQEMVHLVNKESSETPDQFMTADETNLQNVDM 489  
 DB 717 KKKLVINGANGAVDRKPSGNGEASKSQEMVHLVNKESSETPDQFMTADETNLQNVDM 776  
 QY 490 KIGV 493  
 DB 777 KIGV 780

## RESULT 13

097569 PRELIMINARY; PRT; 364 AA.  
 ID 097569;  
 AC 097569;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE CD44 (Fragment).  
 OS Ceratotherium simum simum.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Rhinocerotidae; Ceratotherium.  
 NCBI\_TaxID=73337;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Worley M.B.;  
 RT "White rhinoceros CD44."  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF045933; AAD02418.1; -.  
 DR HSSP; P98066; ITRG.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.  
 DR GO; GO:0005540; F:hyaluronic acid binding; IEA.  
 DR GO; GO:0007155; P:cell adhesion; IEA.  
 DR InterPro; IPR001231; CD44\_antigen.  
 DR InterPro; IPR000538; Link.  
 DR Pfam; PF00193; Xlink; 1.  
 DR PRINTS; PR00658; CD44.  
 DR PRINTS; PR01265; LINKMODULE.  
 DR ProDom; PD000918; Link; 1.  
 DR SMART; SM00445; Link; 1.  
 DR PROSITE; PS01241; Link; 1.  
 DR NON\_TER 364  
 FT SEQUENCE 364 AA; 39670 MW; C56B5E410AABICFP CRC64;

Query Match 58.9%; Score 1536.5; DB 6; Length 364;  
 Best Local Similarity 64.1%; Pred. No. 3.6e-101;  
 Matches 318; Conservative 17; Mismatches 26; Indels 135; Gaps 8;

QY 1 MDKFWMAWG-LCLVPLSLA--QIDNITCRFAGVHVEKNRGYSISRTAADLCKAFN 60  
 DB 1 MDKFWMAWG-LCLVPLSLA--QIDNITCRFAGVHVEKNRGYSISRTAADLCKAFN 60  
 QY 61 PTMAQEKALISGFETCRGFIIEGHVIVPRHPNSICANNVTGYI-LTNTSGOYDTYCN 120  
 DB 61 PTMAQEKALISGFETCRGFIIEGHVIVPRHPNSICANNVTGYI-LTNTSGOYDTYCN 120  
 QY 121 ASAPP-EEDCTSVTDLPNAPFDGPIITITVNRDGRVYQKGEYRTNPEIDYPSNFTD 179  
 DB 121 ASAPP-EEDCTSVTDLPNAPFDGPIITITVNRDGRVYQKGEYRTNPEIDYPSNFTD 180  
 QY 180 SCSSSRSSSTSGGY-IFYT-FSTVHPIDEDSPWTTDSTRIPATNDSHSTLLQPTAN 237  
 DB 181 SCSSSRSSSTSGGY-IFYT-FSTVHPIDEDSPWTTDSTRIPATNDSHSTLLQPTAN 237  
 QY 238 PNTGLVEDLDRGTPLSMTTQGSNSQSFISHGIEEDKDPHTTSTLTSNRNDVYTGRRD 297  
 DB 217 ----- 216

QY 298 PNHSEGTTLLEGYTHSHVPHTKESRTIPVTSKATGSGVAVTVGDSNSNVRSLSGQ 357  
DB 217 --HSENT-----PTTKDQ-----GSDPH----- 233  
QY 358 DTFPSGSHHTHGESDGHSHGSGEGANTSGPIRTPOIPEMLIILASLALAILAV 417  
DB 234 -----SGRSHHTHGTESPGVSRSGEGANTSGPIRKPPIPEMLIILASLALAILAV 288  
QY 418 CIANVSRRCGQKKCLVINGNGAVEDRKPSGLNGEASKSOEVLVNNESSTPDQFMT 477  
DB 289 CIANVSRRCGQKKCLVINGNGAVEDRKASGLNGEASKSOEVLVNNESSTPDQFMT 348  
QY 478 ADETNPLOVDMKIGV 493  
DB 349 ADETNPLOVDMKIGV 364

## RESULT 14

Q92493 PRELIMINARY; PRT; 294 AA.  
AC Q92493  
DT 01-FEB-1997 (TREMblrel. 02, Created)  
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)  
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
DE Cell surface glycoprotein CD44.  
GN CD44.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lymphoblast;  
RX MEDLINE=92017823; PubMed=1922057;  
RA Shrivvelman E., Bishop J.M.;  
RT "Expression of CD44 is repressed in neuroblastoma cells.";  
RL Mol. Cell. Biol. 11:5446-5453(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lymphoblast;  
RX MEDLINE=93101687; PubMed=1465456;  
RA Screation G.R., Bell M.V., Jackson D.G., Cornells F.B., Gerth U.,  
RA Bell J.I.;  
RT "Genomic structure of DNA encoding the lymphocyte homing receptor CD44  
RT reveals at least 12 alternatively spliced exons.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:12160-12164(1992).  
DR EMBL; L05423; AAB13622.1; -.  
DR EMBL; M69215; AAB13622.1; JOINED.  
DR EMBL; L05407; AAB13622.1; JOINED.  
DR EMBL; L05408; AAB13622.1; JOINED.  
DR EMBL; L05410; AAB13622.1; JOINED.  
DR EMBL; L05410; AAB13622.1; JOINED.  
DR EMBL; L05420; AAB13622.1; JOINED.  
DR EMBL; L05421; AAB13622.1; JOINED.  
DR EMBL; L05422; AAB13622.1; JOINED.  
DR PIR; JH0417; JH0417.  
DR HSSP; P98066; ITSG.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.  
DR GO; GO:0005540; F:hyaluronic acid binding; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR001231; CD44\_antigen.  
DR InterPro; IPR000538; Link.  
DR Pfam; PF00193; Xlink; 1.  
DR PRINTS; PR00658; CD44.  
DR PRINTS; PR01265; LINKMODULE.  
DR PRODOM; PD000918; Link; 1.  
DR SMART; SM00445; Link; 1.  
DR PROSITE; PS01241; Link; 1.  
SQ SEQUENCE 294 AA; 32074 MW; E11B9B38FA74817 CRC64;

Query Match 57.3%; Score 1494; DB 4; Length 294;  
Best Local Similarity 68.7%; Pred. No. 2.8e-98;

Matches 292; Conservative 0; Mismatches 1; Indels 132; Gaps 1;  
QY 1 MDKFWMAAMGCLVPLSLAQLDINTCRFAGVHVENKGRYSISRTAADLCAFNSTL 60  
DB 1 MDKFWMAAMGCLVPLSLAQLDINTCRFAGVHVENKGRYSISRTAADLCAFNSTL 60  
QY 61 PTMAQMEKALISGRTCYGFIIEGVVYIPRIHPNSICAAANTGYIILTSNTSOQDTYCFN 120  
DB 61 PTMAQMEKALISGRTCYGFIIEGVVYIPRIHPNSICAAANTGYIILTSNTSOQDTYCFN 120  
QY 121 ASAPPEEDCTSVTLPLNAPDGPITITIVNRDGTYYVQGEYRTNPEDIPENPTDDVSS 180  
DB 121 ASAPPEEDCTSVTLPLNAPDGPITITIVNRDGTYYVQGEYRTNPEDIPENPTDDVSS 180  
QY 181 GSSSERSTSGGYIFTFSTYVHPIDEDSPWITDSTDRIPATNMDSHSTTLQPTANPNT 240  
DB 181 GSSSERSTSGGYIFTFSTYVHPIDEDSPWITDSTDRIPATNMDSHSTTLQPTANPNT 223  
QY 241 GLVEDLDRGTPLSMWTQGSNSGFSSTSGHEGLEDKDHPTTSTLTSNRRNDVYGRDPNH 300  
DB 224 ----- 223  
QY 301 SEGSTTLLEGYTHSHVPHTKESRTIPVTSKATGSGVAVTVGDSNSNVRSLSGQDPTF 360  
DB 224 -----DQDTF 228  
QY 361 HPSGSHHTHGESDGHSHGSGEGANTSGPIRTPOIPEMLIILASLALAILAV 420  
DB 229 HPSGSHHTHGESDGHSHGSGEGANTSGPIRTPOIPEMLIILASLALAILAV 288  
QY 421 VNSRR 425  
DB 289 VNSRR 293

## RESULT 15

Q70509 PRELIMINARY; PRT; 364 AA.  
AC Q70509;  
DT 01-AUG-1998 (TREMblrel. 07, Created)  
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
DE Glycoprotein CD44s (Rat).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Lumbar spine;  
RA Stevens J.W.;  
RT Submitted (May-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF065147; AAC17117.1; -.  
DR HSSP; P98066; ITSG.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.  
DR GO; GO:0005540; F:hyaluronic acid binding; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR001231; CD44\_antigen.  
DR InterPro; IPR000538; Link.  
DR Pfam; PF00193; Xlink; 1.  
DR PRINTS; PR00658; CD44.  
DR PRINTS; PR01265; LINKMODULE.  
DR PRODOM; PD000918; Link; 1.  
DR SMART; SM00445; Link; 1.  
DR PROSITE; PS01241; Link; 1.  
SQ SEQUENCE 364 AA; 39725 MW; BA249776C4419AA7 CRC64;

Query Match 52.4%; Score 1367.5; DB 11; Length 364;  
Best Local Similarity 59.2%; Pred. No. 3.6e-89;  
Matches 295; Conservative 20; Mismatches 44; Indels 139; Gaps 10;  
QY 1 MDKFWMAAMGCLVPLSLA--QIDINTCRFAGVHVENKGRYSISRTAADLCAFN 57

```
Db      1 MDKVMHTAWGLCLLQSLAQOQIDINTICRYAGFHEKNGRYSISRTAEADLCEAFN 60
QY      58 STLPYMAQMEKALISIGETCRYGFIIEGHVVIPIRIHNSICAANTGVYI-LTSNTSOYDT 116
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 TLPYMAQMEKALISIGETCRYGFIIEGHVVIPIRIHNAICAAANTGVYIILASNTSHYDT 120
QY      117 YCFNAPAPPEEDCTSTVDLPNAPDGPITTTIVNRDGRIVQKGEYRINPEDIYPSNYPTD 176
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121 YCFNAPAPPEEDCTSTVDLPNAPDGPITTTIVNRDGRIVQKGEYRINPEDIYPSNYPTD 180
QY      177 DVSSGSSSRSTSGYIFFTFSTVHPIDEDSPWITDSTDRIPATNMDSHSTTLOPTA 236
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      181 DVSSGSTIEK-STEGYIILHT-----DLP-----TSQPT- 208
QY      237 NPNTGLVEDLDRIQPLSMTTQGSNSQSFSSTHEGLEEDKDHPPTSTLTSSNRNDVTGRR 296
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      209 -----GDR 211
QY      297 DPNHSEGSTTILEGYTSHYPHTKESRTFIPVTSAKTGSFGVTAVTVGDSNVNRSLSGD 356
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      212 DDAFFIGST-----LATSDGSSMD----- 231
QY      357 QDTFHPGGSHT-THGSESDGSHSGQEGGANTSGPIRTPOIPEWLIILASILALALIL 415
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      232 ----PRGFPDVTWHSBELAGHSSGNQDSGVTTSQPARRPQIPEWLIILASILALALIL 286
QY      416 AVCIAVNSRRRCQKKLVINSNGAVEDRKPSPGLNGEASKSOEMVHLVNXESSETPDOF 475
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      287 AVCIAVNSRRRCQKKLVINSNGAVEDRKPSPGLNGEASKSOEMVHLVNXESSETPDOF 346
QY      476 MTADETRNLOVDMKIGV 493
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      347 MTADETRNLOVDMKIGV 364
```

Search completed: March 8, 2004, 06:11:38  
Job time : 50 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 8, 2004, 06:08:16 ; Search time 59 Seconds  
(without alignments)  
2360,948 Million cell updates/sec

Title: PCT-US01-51014-1

Perfect score: 2608

Sequence: 1 MDKFWMHAAWGICLVPLSLA.....QPMADETRNLDQNMKIGV 493

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: \_Geneseq\_290a04:\*  
2: geneseqp1980s:\*  
3: geneseqp1990s:\*  
4: geneseqp2000s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003bs:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2608	100.0	493	5	AAU99123	AAU99123 Haematopo
2	2608	100.0	493	6	ABU04623	ABU04623 Human exp
3	2608	100.0	493	6	ABU04612	ABU04612 Human exp
4	2608	99.3	493	2	AAK20817	AAK20817 Epithelia
5	2608	99.3	493	2	AAW80454	AAW80454 Human CD4
6	2608	99.3	493	2	AAW89151	AAW89151 Human CD4
7	2608	99.3	493	3	AAV96140	AAV96140 Human epi
8	2608	99.3	493	4	AAU02449	AAU02449 Human epi
9	2608	99.3	493	6	ABU04637	ABU04637 Human exp
10	2608	99.3	493	6	ABU04627	ABU04627 Human exp
11	2608	99.3	493	6	ABU04639	ABU04639 Human exp
12	2608	99.3	493	6	ABU04633	ABU04633 Human exp
13	2608	99.3	493	6	ABU04635	ABU04635 Human exp
14	2608	99.3	493	6	ABU04613	ABU04613 Human exp
15	2608	99.0	493	2	AAK91445	AAK91445 Human epi
16	2608	99.0	493	6	ABU04631	ABU04631 Human exp
17	2608	99.0	493	6	ABU04621	ABU04621 Human exp
18	2608	99.5	699	6	ABU04670	ABU04670 Lung can
19	2608	99.5	699	6	ABU04647	ABU04647 Human exp
20	2608	99.5	699	6	ABU04614	ABU04614 Human exp
21	2608	99.5	699	6	ABU04608	ABU04608 Human exp
22	2608	99.5	699	7	ADD90586	ADD90586 Human CD4
23	2608	99.5	700	4	AAV97579	AAV97579 Human CD4
24	2608	99.5	700	6	ABU04640	ABU04640 Human exp
25	2608	99.5	700	7	ADD90594	ADD90594 Human CD4

26	2485	95.3	742	4	ABG17071	ABG17071 Novel hum
27	2468.5	94.7	742	6	ABU04620	ABU04620 Human exp
28	2468.5	94.7	742	6	ABU04645	ABU04645 Human exp
29	2466	94.6	470	6	ABU04651	ABU04651 Human exp
30	2466	94.6	470	6	ABU04603	ABU04603 Human exp
31	2457.5	94.2	742	6	ABU04653	ABU04653 Human exp
32	2457.5	94.2	742	6	ABU04616	ABU04616 Human exp
33	2353.5	90.2	668	6	ABU04619	ABU04619 Human exp
34	2348	90.0	676	6	ABU04602	ABU04602 Human exp
35	2348	90.0	676	6	ABU04652	ABU04652 Human exp
36	2326.5	89.2	719	6	ABU04604	ABU04604 Human exp
37	2326.5	89.2	719	6	ABU04650	ABU04650 Human exp
38	2203	84.5	625	6	ABU04617	ABU04617 Human exp
39	2121.5	81.3	675	6	ABU04618	ABU04618 Human exp
40	2037.5	78.1	608	6	ABP72424	ABP72424 Human CD4
41	2036	78.1	395	6	ABU04615	ABU04615 Human exp
42	1841	70.6	361	6	ABU04622	ABU04622 Human exp
43	1841	70.6	361	6	AAE30338	AAE30338 Human CD4
44	1841	70.6	361	7	ADD90592	ADD90592 Human CD4
45	1838	70.5	361	6	ABU04607	ABU04607 Human exp

#### ALIGNMENTS

RESULT 1  
AAU99123 standard; protein; 493 AA.

AAU99123; 24-SEP-2002 (first entry)  
Hematopoietic cell E-selectin/L-selectin ligand (HCELU) protein.  
HCELU, hematopoietic cell E-selectin/L-selectin ligand (HCELU) protein.  
hematopoietic progenitor cell; stem cell isolation; engraftment; human;  
hematopoietic disorder; leukemia; inflammatory disorder; stroke;  
myocardial infarction; Parkinson's disease; diabetes; muscle dystrophy;  
liver disorder; rheumatoid arthritis; inflammatory bowel disease.  
Homo sapiens.  
W0200244342-A2.  
06-JUN-2002.  
18-OCT-2001; 2001WO-US051014.  
18-OCT-2000; 2000US-0240987P.  
11-JUN-2001; 2001US-0297474P.  
(BGM) BRIGHAM & WOMENS HOSPITAL INC.  
Sackstein R;  
WPI; 2002-527707/56.  
Claim 7; Page 10; 94pp; English.  
This invention relates to a purified glycosylated polypeptide expressed on normal human hematopoietic progenitor cells and on leukemic blasts, termed hematopoietic cell E-selectin/L-selectin ligand (HCELU). The HCELU protein of the invention is useful for identifying a stem cell, for isolating a stem cell from a population of cells or for increasing the engraftment potential of a cell population. The protein is useful for treating a hematopoietic disorder such as leukemia, or inflammatory disorder in a subject. Methods of the invention are useful for treating a hematopoietic disorder in a subject or mammal, and treating cancer or a

disorder amenable for treatment with a stem cell in a subject or mammal, where the disorder is selected from myocardial infarction, Parkinson's disease, diabetes, congenital muscle dystrophy, stroke, genetic/congenital disorders and liver disorders. Another method disclosed in the invention is useful for increasing levels of engrafted stem cells in a subject, preferably a human suffering from or at risk for hematopoietic disorder or cancer, preferably blood cancer, respectively. The HCEIL protein is useful for treating hematopoietic disorders such as aplastic anaemia, and inflammatory disorders such as rheumatoid arthritis, inflammatory bowel disease and asthma. The HCEIL protein is useful as an immunogen for producing anti-HCEIL antibodies. The present sequence represents the HCEIL protein of the invention

Sequence 493 AA:

Query Match 100.0%; Score 2608; DB 5; Length 493;  
Best Local Similarity 100.0%; Pred. No. 1.2e-191;  
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRFWMAWGLCLVPLSLAQIDINICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60  
DB 1 MDRFWMAWGLCLVPLSLAQIDINICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60  
QY 61 PTMAQMEKALSIGETCRVGFIEGHVYIPRIHNSICANNNGVYILTSNTSQDYTCFN 120  
DB 61 PTMAQMEKALSIGETCRVGFIEGHVYIPRIHNSICANNNGVYILTSNTSQDYTCFN 120  
QY 121 ASAPPEEDCTSVTDLPNAPDPIITITIVNRDGTQRYVQKGEYRINPEDIYPSNPTDDVSS 180  
DB 121 ASAPPEEDCTSVTDLPNAPDPIITITIVNRDGTQRYVQKGEYRINPEDIYPSNPTDDVSS 180  
QY 181 GSSSERSSTSGGYIFTFSTYVHPIDEDSPWIDSDIRIPATNMDSHSTLQPTANPT 240  
DB 181 GSSSERSSTSGGYIFTFSTYVHPIDEDSPWIDSDIRIPATNMDSHSTLQPTANPT 240  
QY 241 GLVEDLDRTGFLPMTQOSNSQSFSTSHGLEGEDKHPITSTLSSNRNDVTGGRDPNH 300  
DB 241 GLVEDLDRTGFLPMTQOSNSQSFSTSHGLEGEDKHPITSTLSSNRNDVTGGRDPNH 300  
QY 301 SEGSTTLEGTSHYPTKESRPTIPYTSAKTSFGVTATVGDSDSNVNRSLSGQDPT 360  
DB 301 SEGSTTLEGTSHYPTKESRPTIPYTSAKTSFGVTATVGDSDSNVNRSLSGQDPT 360  
QY 361 HPSGSHTTGSESDGSHSQSGEGGANTSGPIRTPOIPMILITLILALAILAVCIA 420  
DB 361 HPSGSHTTGSESDGSHSQSGEGGANTSGPIRTPOIPMILITLILALAILAVCIA 420  
QY 421 VNSRRRCGQKKKLYINGNGAVEDRKPGLNGEASKSQEVAHLVNRKSSSTPQCFNTAB 480  
DB 421 VNSRRRCGQKKKLYINGNGAVEDRKPGLNGEASKSQEVAHLVNRKSSSTPQCFNTAB 480  
QY 481 TRNLQNDMKIGV 493  
DB 481 TRNLQNDMKIGV 493

RESULT 2

ID ABU04623 standard; protein; 493 AA.

AC ABU04623;

DT 29-JAN-2003 (first entry)

DE Human expressed protein tag (EPT) #1289.

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;

KW protease; protease inhibitor; transporter; cytoskeletal protein;

KW receptor; transcription factor; cancer; MHC;

KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;

XX adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.

XX WO200278524-A2.  
PN 10-OCT-2002.  
XX 28-MAR-2002; 2002WO-US009671.  
XX 28-MAR-2001; 2001US-0279495P.  
XX 21-MAY-2001; 2001US-0292544P.  
XX 08-AUG-2001; 2001US-0310801P.  
XX 01-OCT-2001; 2001US-03262370P.  
XX 04-DEC-2001; 2001US-0356780P.  
XX 20-FEB-2002; 2002US-0358985P.  
XX (ZYCO-) ZYCO8 INC.  
XX Chicx RM, Tomlinson AJ, Urban RG;  
XX WPI; 2003-040607/03.  
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
XX cytoskeletal proteins, receptors or transcription factors), useful for  
XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
XX leukemia.  
XX Example 2; SEQ ID NO 1289; 134pp; English.

The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling. Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 493 AA:

Query Match 100.0%; Score 2608; DB 6; Length 493;

Best Local Similarity 100.0%; Pred. No. 1.2e-191; Indels 0; Gaps 0;

Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRFWMAWGLCLVPLSLAQIDINICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60

DB 1 MDRFWMAWGLCLVPLSLAQIDINICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60

QY 61 PTMAQMEKALSIGETCRVGFIEGHVYIPRIHNSICANNNGVYILTSNTSQDYTCFN 120

DB 61 PTMAQMEKALSIGETCRVGFIEGHVYIPRIHNSICANNNGVYILTSNTSQDYTCFN 120

QY 121 ASAPPEEDCTSVTDLPNAPDPIITITIVNRDGTQRYVQKGEYRINPEDIYPSNPTDDVSS 180

DB 121 ASAPPEEDCTSVTDLPNAPDPIITITIVNRDGTQRYVQKGEYRINPEDIYPSNPTDDVSS 180

QY 181 GSSSERSSTSGGYIFTFSTYVHPIDEDSPWIDSDIRIPATNMDSHSTLQPTANPT 240

DB 181 GSSSERSSTSGGYIFTFSTYVHPIDEDSPWIDSDIRIPATNMDSHSTLQPTANPT 240

QY 241 GLVEDLDRTGFLPMTQOSNSQSFSTSHGLEGEDKHPITSTLSSNRNDVTGGRDPNH 300

DB 241 GLVEDLDRTGFLPMTQOSNSQSFSTSHGLEGEDKHPITSTLSSNRNDVTGGRDPNH 300

QY 301 SEGSTTLEGTSHYPTKESRPTIPYTSAKTSFGVTATVGDSDSNVNRSLSGQDPT 360

Db 301 SEGSTTLLEGYTHSHYPRHKESRTFIPVTSKATGSGVTAIVYGDNSNVNRSLSGDDTF 360  
 QY 361 HPSGGSHHTHGSSEDSHSGSQEGANTTSGPIRTPOIPEWLIILASLALAILAVCIA 420  
 Db 361 HPSGGSHHTHGSSEDSHSGSQEGANTTSGPIRTPOIPEWLIILASLALAILAVCIA 420  
 QY 421 VNSRRRCQKKLVINSNGAVEDRRKPSGLNGEASKSQEMVHLVKNESSETPDQMTADE 480  
 Db 421 VNSRRRCQKKLVINSNGAVEDRRKPSGLNGEASKSQEMVHLVKNESSETPDQMTADE 480  
 QY 481 TRNLQNVDMKIGV 493  
 Db 481 TRNLQNVDMKIGV 493  
 RESULT 3  
 AB004612  
 ID AB004612 standard; protein; 493 AA.  
 AC AB004612;  
 XX  
 XX 29-JAN-2003 (first entry)  
 DE Human expressed protein tag (EPT) #1278.  
 XX  
 KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
 KW protease; protease inhibitor; transporter; cytoskeletal protein;  
 KW receptor; transcription factor; cancer; MHC;  
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.  
 OS  
 XX Homo sapiens.  
 PN WO200278524-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 28-MAR-2002; 2002MO-US009671.  
 XX  
 PR 28-MAR-2001; 2001US-0279495P.  
 PR 21-MAY-2001; 2001US-0292544P.  
 PR 08-AUG-2001; 2001US-0310801P.  
 PR 01-OCT-2001; 2001US-0326370P.  
 PR 04-DEC-2001; 2001US-0336780P.  
 PR 20-FEB-2002; 2002US-0358985P.  
 PR  
 PA (ZYCO-) ZYCOs INC.  
 PA  
 PI Chicz RM, Tomlinson AJ, Urban RG;  
 PI  
 XX WPI; 2003-040607/03.  
 DR  
 XX  
 XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.  
 XX  
 PS Example 2; SEQ ID NO 1278; 134pp; English.  
 CC The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukaemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational

CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIP0 at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 493 AA;  
 Query Match 100.0%; Score 2608; DB 6; Length 493;  
 Best Local Similarity 100.0%; Pred. No.1,2e-191;  
 Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MDKFWHAAAGCLVPLSLAQIDNLITCRPAGVHYEKNGRYSISRTAADLCKANSTL 60  
 Db 1 MDKFWHAAAGCLVPLSLAQIDNLITCRPAGVHYEKNGRYSISRTAADLCKANSTL 60  
 QY 61 PTMAQMEKALISGFETCRYGFTGSHVIRIHNSICAANNQGVYILTSNTSOYPTCYCFN 120  
 Db 61 PTMAQMEKALISGFETCRYGFTGSHVIRIHNSICAANNQGVYILTSNTSOYPTCYCFN 120  
 QY 121 ASAPBEDCTSVTDLPNAPDGPITITIVNRDGRVYQKGEYRTNPEDIYPSNPTDVS 180  
 Db 121 ASAPBEDCTSVTDLPNAPDGPITITIVNRDGRVYQKGEYRTNPEDIYPSNPTDVS 180  
 QY 181 GSSSESSSTSGGYTFYFTSTVHPIDEDSPWITDSTDRIPATNMDSSHSTTLQPTANPT 240  
 Db 181 GSSSESSSTSGGYTFYFTSTVHPIDEDSPWITDSTDRIPATNMDSSHSTTLQPTANPT 240  
 QY 241 GLVEDLDRTGPLEMTTQOSNSQSFSTSHGLBEDKDHPTTSTLSSNRNDVTGRRDPNH 300  
 Db 241 GLVEDLDRTGPLEMTTQOSNSQSFSTSHGLBEDKDHPTTSTLSSNRNDVTGRRDPNH 300  
 QY 301 SEGSTTLLEGYTHSHYPRHKESRTFIPVTSKATGSGVTAIVYGDNSNVNRSLSGDDTF 360  
 Db 301 SEGSTTLLEGYTHSHYPRHKESRTFIPVTSKATGSGVTAIVYGDNSNVNRSLSGDDTF 360  
 QY 361 HPSGGSHHTHGSSEDSHSGSQEGANTTSGPIRTPOIPEWLIILASLALAILAVCIA 420  
 Db 361 HPSGGSHHTHGSSEDSHSGSQEGANTTSGPIRTPOIPEWLIILASLALAILAVCIA 420  
 QY 421 VNSRRRCQKKLVINSNGAVEDRRKPSGLNGEASKSQEMVHLVKNESSETPDQMTADE 480  
 Db 421 VNSRRRCQKKLVINSNGAVEDRRKPSGLNGEASKSQEMVHLVKNESSETPDQMTADE 480  
 QY 481 TRNLQNVDMKIGV 493  
 Db 481 TRNLQNVDMKIGV 493  
 RESULT 4  
 AAR20817  
 ID AAR20817 standard; protein; 493 AA.  
 XX  
 AC AAR20817;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT 21-MAY-1992 (first entry)  
 XX  
 DE Epithelial CD44 Antigen.  
 XX  
 KW Rapid immunoselection cloning technique; cell surface antigen;  
 KW epithelium; tumour invasiveness; carcinoma.  
 OS  
 XX Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT Peptide 1..20  
 FT Domain /label= signal  
 FT 21..400  
 FT Modified-site /label= extracellular  
 FT 65..67  
 FT /label= N-linked glycosylation  
 FT /note= "putative"  
 FT Modified-site 97..99  
 FT /label= N-linked glycosylation

```

FT      /note= "putative"
FT      140..142
FT      Modified-site
FT      /label= N-linked glycosylation
FT      /note= "putative"
FT      150..152
FT      Modified-site
FT      /label= N-linked glycosylation
FT      /note= "putative"
FT      160..162
FT      Modified-site
FT      /label= N-linked glycosylation
FT      /note= "putative"
FT      180..181
FT      Modified-site
FT      /label= glycosylation
FT      /note= "potential site of attachment of serine-linked
FT      chondroitin sulphate"
FT      190..191
FT      Modified-site
FT      /label= glycosylation
FT      /note= "potential site of attachment of serine-linked
FT      chondroitin sulphate"
FT      295..297
FT      Modified-site
FT      /label= N-linked glycosylation
FT      /note= "putative"
FT      390..391
FT      Modified-site
FT      /label= glycosylation
FT      /note= "potential site of attachment of serine-linked
FT      chondroitin sulphate"
FT      401..421
FT      Region
FT      /label= transmembrane
FT      422..493
FT      Domain
FT      /label= cytoplasmic

PN      W09201049-A.
XX      23-JAN-1992.
PD      13-JUL-1990; 90US-00553759.
XX      13-JUL-1990; 90US-00553759.
PR      13-JUL-1990; 90US-00553759.
XX      (GENO ) GEN HOSPITAL CORP.
PA      Seed B, Aruffo A, Amiot M;
PI      WPI, 1992-056864/07.
XX      N-PSDB; AAQ21186.
DR      N-PSDB; AAQ21186.
XX      New CD53 cell surface antigen and DNA encoding it - for immuno-therapy
PT      and diagnosis of haematopoietic neoplasms, etc.
XX      Example 15; Page 117; 160pp; English.
PS      The cDNA encoding epithelial CD44 is quite similar to haematopoietic
XX      CD44.5 cDNA (see AAQ21185), but encodes an additional extracellular
CC      domain of 165 amino acids, inserted about 140 residues upstream of the
CC      transmembrane section shared by both clones. The ability to interfere
CC      with binding of epithelial CD44 to extracellular matrices can be useful
CC      in diagnosis and therapy. e.g. to diminish the likelihood of metastasis
CC      in cancer patients. Soluble forms of CD44 can act to prevent metastatic
CC      cells from "homing" to lymph nodes. (Updated on 25-MAR-2003 to correct PA
CC      field.)
XX      Sequence 493 AA;
SQ

```

```

Query Match          99.3%; Score 2590; DB 2; Length 493;
Best Local Similarity 99.4%; Pred. No. 2,9e-190;
Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 MDKFWMAWGLCVPLSLAQLDINTCFRAGVHVEKNGRYSISRTAADLCAFNSTL 60
DB      1 MDKFWMAWGLCVPLSLAQLDINTCFRAGVHVEKNGRYSISRTAADLCAFNSTL 60
QY      61 PTMAQEKALISGFETCRYGFTGEGHVVPRIHPSICAAANTGVYIITSNTSOYDYCFN 120
DB      61 PTMAQEKALISGFETCRYGFTGEGHVVPRIHPSICAAANTGVYIITSNTSOYDYCFN 120

```

```

QY      121 ASAPPEEDCTSVTDLPNAPDGPITITTVNRDGRYVOKGEYRNPEDIPYENPTDDVSS 180
DB      121 ASAPPEEDCTSVTDLPNAPDGPITITTVNRDGRYVOKGEYRNPEDIPYENPTDDVSS 180
QY      181 GSSSERSSSTGGGYFYNESTVHPIDPDSPMITSTDRIPATNNDSSHTLQPTANPT 240
DB      181 GSSSERSSSTGGGYFYNESTVHPIDPDSPMITSTDRIPATNNDSSHTLQPTANPT 240
QY      241 GLVEDLDRTGPLSMITQQSNSQSFSTSHGLEEDKDPHTSTLTSNRNDVTGGRDPNH 300
DB      241 GLVEDLDRTGPLSMITQQSNSQSFSTSHGLEEDKDPHTSTLTSNRNDVTGGRDPNH 300
QY      301 SEGSTTLLEGTSYHPHTEKERTPIPTSAATGSGFYTAAYVGDPSNVNRSLSGDDPTF 360
DB      301 SEGSTTLLEGTSYHPHTEKERTPIPTSAATGSGFYTAAYVGDPSNVNRSLSGDDPTF 360
QY      361 HPSGGSHHTTGSSESDSHSGSQEGANTTSGPIRTPOIPEWLLILASILALAILAVCIA 420
DB      361 HPSGGSHHTTGSSESDSHSGSQEGANTTSGPIRTPOIPEWLLILASILALAILAVCIA 420
QY      421 VNSRRRCQKKLVINGNGAVEDRKSGCLNGEASKCEMVLNKSSETPDQFTADE 480
DB      421 VNSRRRCQKKLVINGNGAVEDRKSGCLNGEASKCEMVLNKSSETPDQFTADE 480
QY      481 TRNLQNVDMKIGV 493
DB      481 TRNLQNVDMKIGV 493

```

```

RESULT 5
AAW80454
ID      AAW80454 standard; protein; 493 AA.
XX
AC      AAW80454;
XX
DT      25-MAR-2003 (revised)
DT      07-JUN-1999 (first entry)
XX
DE      Human CD44 antigen (epithelial form).
XX
KW      CD44; cell surface antigen; human; cDNA library; T lymphocyte;
KW      metastasis; therapy.
XX
OS      Homo sapiens.
XX
FH      Key
FT      Location/Qualifiers
FT      Domain
FT      /note= "transmembrane domain"
XX
PN      US5830731-A.
XX
PD      03-NOV-1998.
XX
PF      21-MAY-1997; 97US-00861205.
XX
PR      25-FEB-1988; 88US-00160416.
PR      13-JUL-1989; 89US-00379076.
PR      23-MAR-1990; 90US-00498809.
PR      13-JUL-1990; 90US-00553759.
PR      01-DEC-1992; 92US-00983647.
XX
PA      (GENO ) GEN HOSPITAL CORP.
PI      Seed B, Aruffo A;
DB      WPI, 1998-609251/51.
DB      N-PSDB; AAV63462.
PT      New cloning vector and polylinker - based on existing sequences for
PT      efficient cloning and expression of mammalian cDNA(s), especially human
PT      lymphocyte antigenic sequences.
XX

```



PS Example 15; Col 73-76; 75pp; English.

XX This polypeptide comprises human CD44 antigen. Its amino acid sequence  
CC was deduced from the nucleotide sequence (see AAV63462) of a cDNA clone  
CC isolated from a colon carcinoma HT29 cell library using a novel method  
CC for cloning cDNAs from mammalian expression libraries. The method is  
CC based on transient expression of an antigen in eukaryotic cells and  
CC physical selection of cells expressing the antigen by adhesion to an  
CC antibody-coated substrate. It is useful for the isolation and cloning of  
CC any protein which can be expressed and transported to the cell surface  
CC membrane of a eukaryotic cell, and was used to clone genes (see AAV63442-  
CC 63) encoding cell surface antigens from mammalian lymphocytes (see  
CC AAV60440-55). The purified genes and proteins are useful for  
CC immunodiagnostic and immunotherapeutic applications, including the  
CC diagnosis and treatment of immune-mediated infections, diseases, and  
CC disorders of animals, including humans. The epithelial form of CD44 is a  
CC glycosylated protein of about 160 kDa. The ability to interfere with the  
CC binding of epithelial CD44 with extracellular matrices may be useful in  
CC therapy or diagnosis, e.g. to diminish the likelihood of metastasis in  
CC cancer patients. Soluble forms of CD44 can act to prevent metastatic  
CC cells from homing to lymph nodes. (Updated on 25-MAR-2003 to correct PR  
CC field.)

XX Sequence 493 AA;

XX Query Match

99.3%; Score 2590; DB 2; Length 493;

Best Local Similarity 99.4%; Pred. No. 2.9e-190; Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKFWHAAWGLCLVPLSLAQLDNLITCRFAGVHVEKNGRYSISRTBADLCKAFNSTL 60  
DB 1 MDKFWHAAWGLCLVPLSLAQLDNLITCRFAGVHVEKNGRYSISRTBADLCKAFNSTL 60  
QY 61 PTMAQMEKALSIGETCRGFIHGVIPRIHPNSICANNVGYLLVNTSQTDTYCFN 120  
DB 61 PTMAQMEKALSIGETCRGFIHGVIPRIHPNSICANNVGYLLVNTSQTDTYCFN 120  
QY 121 ASAPPEEDCTSVTDLPAFADGPIITITVNRDGTTRYVQGEYRTNPEDIYPSNPTDDVSS 180  
DB 121 ASAPPEEDCTSVTDLPAFADGPIITITVNRDGTTRYVQGEYRTNPEDIYPSNPTDDVSS 180  
QY 181 GSSSERSTSGGYIFTFSTVHPPIEDSDPWITDSIDRIATMDSHSTLLOPTANPT 240  
DB 181 GSSSERSTSGGYIFTFSTVHPPIEDSDPWITDSIDRIATMDSHSTLLOPTANPT 240  
QY 241 GLVVDLDRGTPLSMITTOQNSQSFSTSHGLEEDKHPITSTLSSNRNDVTCGRADPNH 300  
DB 241 GLVVDLDRGTPLSMITTOQNSQSFSTSHGLEEDKHPITSTLSSNRNDVTCGRADPNH 300  
QY 301 SEGSTTLLEGYTSHPPTKESRTFIPVTSKATGSGVTAIVGDSNNSNVRSLSGQDTF 360  
DB 301 SEGSTTLLEGYTSHPPTKESRTFIPVTSKATGSGVTAIVGDSNNSNVRSLSGQDTF 360  
QY 361 HPSSGSHTTGSESDGSHSGOEGAGNTSGPIRTPQIPMLIILASLALATILAVCIA 420  
DB 361 HPSSGSHTTGSESDGSHSGOEGAGNTSGPIRTPQIPMLIILASLALATILAVCIA 420  
QY 421 VNSRRRCGQKKKLVINSNGAVEDRKPGLNGEASKSQEVMHLVNKESSTEPQFWTADE 480  
DB 421 VNSRRRCGQKKKLVINSNGAVEDRKPGLNGEASKSQEVMHLVNKESSTEPQFWTADE 480  
QY 481 TRNLQNDMKIGV 493  
DB 481 TRNLQNDMKIGV 493

RESULT 6

AAW89151 standard; protein; 493 AA.

XX AAW89151;  
XX AAW89151;  
DT 10-MAY-1999 (first entry)

XX Human CD44 antigen (epithelial form).

XX CD44; cell surface antigen; human; cDNA library; T lymphocyte;  
KW metastasis; therapy.

XX Homo sapiens.

XX Key Location/Qualifiers  
FT Domain 401..421  
FT /note="transmembrane domain"

XX US5849898-A.

XX 15-DEC-1998.

XX 07-JUN-1995; 95US-00485447.

XX 25-FEB-1988; 88US-00160416.

XX 13-JUN-1989; 89US-00379076.

XX 23-MAR-1990; 90US-00498809.

XX 13-JUN-1990; 90US-00553759.

XX 01-DEC-1992; 92US-00983647.

XX (GENO) GEN HOSPITAL CORP.

XX Seed B, Oquendo C, Camerini D, Stramenkovic I, Stengelin S;  
PI Amiot M, Lauffer L, Allen J, Simons D, Aruffo A;

XX WPI; 1999-069813/06.

XX N-PEDB; AAV81219.

PT cDNA encoding human CD40 antigen - useful for cloning cDNA encoding cell  
PT surface antigens, constructing cDNA libraries, expressing vectors for  
PT expression in eukaryotic cells or their fragments.

PS Example 15; Col 79-82; 79pp; English.

XX This polypeptide comprises human CD44 antigen. Its amino acid sequence  
CC was deduced from the nucleotide sequence (see AAV81219) of a cDNA clone  
CC isolated from a colon carcinoma HT29 cell library using a novel method  
CC for cloning cDNAs from mammalian expression libraries. The method is  
CC based on transient expression of an antigen in eukaryotic cells and  
CC physical selection of cells expressing the antigen by adhesion to an  
CC antibody-coated substrate. It is useful for the isolation and cloning of  
CC any protein which can be expressed and transported to the cell surface  
CC membrane of a eukaryotic cell, and was used to clone genes (see AAV81198-  
CC 220) encoding cell surface antigens such as CD1a, CD1b, CD1c, CD2, CD6,  
CC CD7, CD13, CD14, CD16, CD19, CD20, CD22, CD26, CD27, CD28, CD31, CDW32a,  
CC CDW32b, CD93, CD94, CD96, CD97, CD98, CD99, CD40, CD43, CD44, CD53, ICAM,  
CC LFA-3, FCRIIa, FCRIIb, TILSA and Leu8 (see AAV81198 and AAV8151-52 and  
CC AAV88441). CD40 cDNA (see AAV81198) is specifically claimed. The  
CC epithelial form of CD44 is a glycosylated protein of about 160 kDa. The  
CC ability to interfere with the binding of epithelial CD44 with  
CC extracellular matrices may be useful in therapy or diagnosis, e.g. to  
CC diminish the likelihood of metastasis in cancer patients. Soluble forms  
CC of CD44 can act to prevent metastatic cells from homing to lymph nodes

XX Sequence 493 AA;

XX Query Match

99.3%; Score 2590; DB 2; Length 493;

Best Local Similarity 99.4%; Pred. No. 2.9e-190; Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKFWHAAWGLCLVPLSLAQLDNLITCRFAGVHVEKNGRYSISRTBADLCKAFNSTL 60  
DB 1 MDKFWHAAWGLCLVPLSLAQLDNLITCRFAGVHVEKNGRYSISRTBADLCKAFNSTL 60  
QY 61 PTMAQMEKALSIGETCRGFIHGVIPRIHPNSICANNVGYLLVNTSQTDTYCFN 120  
DB 61 PTMAQMEKALSIGETCRGFIHGVIPRIHPNSICANNVGYLLVNTSQTDTYCFN 120  
QY 121 ASAPPEEDCTSVTDLPAFADGPIITITVNRDGTTRYVQGEYRTNPEDIYPSNPTDDVSS 180

```

Db 121 ASAPBEDCTSVTDLPNAPFDGPIITITVNRDGTTRYVQKGEYRNPEDIPSPNTDDVSS 180
Qy 181 GSSSERSTSGGYFFYFSTVHP1PDEDSPWITDSTRIPATNMDSSHSTTLOPTANPNT 240
Db 181 GSSSERSTSGGYFFYFSTVHP1PDEDSPWITDSTRIPATNMDSSHSTTLOPTANPNT 240
Qy 241 GLVEDLDRTGPIPLSMTTQOQNSQSPSTSHGLEDKDPHTTSTLTSSNRNDVTGRRDPNH 300
Db 241 GLVEDLDRTGPIPLSMTTQOQNSQSPSTSHGLEDKDPHTTSTLTSSNRNDVTGRRDPNH 300
Qy 301 SEGSTTLLEGYTSHPYTKESRTPIPTSAKTSFGVATVAGDSNNVNRSLSGDDPTF 360
Db 301 SEGSTTLLEGYTSHPYTKESRTPIPTSAKTSFGVATVAGDSNNVNRSLSGDDPTF 360
Qy 361 HPSGSHRTTHGSESDGSHSGSQEGGANTTSGPIRTPQIPBWLIIILASLLALAILAVCIA 420
Db 361 HPSGSHRTTHGSESDGSHSGSQEGGANTTSGPIRTPQIPBWLIIILASLLALAILAVCIA 420
Qy 421 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKQEMVHLVNRKESSTPDQFTADE 480
Db 421 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKQEMVHLVNRKESSTPDQFTADE 480
Qy 481 TRNLQNVDMKIGV 493
Db 481 TRNLQNVDMKIGV 493

RESULT 7
AA96140
ID AA96140 standard; protein; 493 AA.
AC AA96140;
XX
DT 19-DEC-2000 (first entry)
XX
DE Human epithelial CD44.
XX
KW CD44.5; cell surface antigen; human; immunoselection; panning;
KW immunodiagnosis; diagnosis; immunotherapy; gene therapy; immune disorder;
KW infection; asthma; immune-complex disease; amyloidosis;
KW multiple sclerosis; parasitic disease; epithelial; metastasis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /label= signal_peptide 20..493
FT /label= Mature_protein 401..421
FT Domain /label= Transmembrane_domain

US611093-A.
XX
PD 29-AUG-2000.
XX
PF 28-OCT-1998; 98US-00181612.
XX
PR 25-FEB-1988; 88US-00160416.
PR 13-JUL-1989; 89US-00379076.
PR 23-MAR-1990; 90US-00498809.
PR 13-JUL-1990; 90US-00553759.
PR 01-DEC-1992; 92US-00983647.
XX
PA (GENO ) GEN HOSPITAL CORP.
PI Stamenkovic I, Seed B;
XX
DR WPI; 2000-586382/55.
DR N-PSDB; AAA50596.
XX
PT Isolated nucleic acid molecule encoding the CD19 cell surface antigen,

```

```

PT useful for immunodiagnosis and immunotherapy of immune-mediated
PT infections or disorders, e.g. asthma, immune-complex disease, parasitic
PT diseases.
PS Example 15; Col 75-78; 75pp; English.
XX
XX The present sequence is that of human epithelial CD44, a glycosylated
CC membrane protein that may promote tumor invasiveness. The sequence was
CC predicted from a cDNA clone (see AAA50598) isolated by screening a cDNA
CC library prepared from the the colon carcinoma line HT29 using the method
CC of the invention. The method, designed to isolate cell surface antigen
CC (CSA) nucleic acids, is based upon transient expression of a CSA in
CC eukaryotic cells and physical selection of cells expressing the antigen
CC by adhesion to (panning on) an antibody-coated substrate such as a
CC culture dish. CSA nucleic acids isolated by the method of the invention,
CC and the proteins they encode, are useful for immunodiagnosis and
CC immunotherapeutic applications, including the diagnosis and treatment of
CC immune-mediated infections, diseases, and disorders in animals, including
CC humans. These disorders include asthma, immune-complex disease,
CC amyloidosis, parasitic diseases or multiple sclerosis. The ability to
CC interfere with the binding of epithelial CD44 with extracellular matrices
CC can be useful in diagnostics and therapy. Interference of the binding can
CC diminish the likelihood of metastasis in cancer patients. Soluble forms
CC of CD44 can act to prevent metastatic cells from homing to lymph nodes
XX
SQ Sequence 493 AA;
Query Match 99.3%; Score 2590; DB 3; Length 493;
Best Local Similarity 99.4%; Pred. No. 2.9e-190;
Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MDKFWMAWGLCLVPLSLAQIDNLICRFAGVHVKNRYSISRTAADLCAFNSTL 60
Db 1 MDKFWMAWGLCLVPLSLAQIDNLICRFAGVHVKNRYSISRTAADLCAFNSTL 60
Qy 61 PTMAOMKALISGFEETRCYGRIBGVVPIRHPNSICANNVGYITLSTSOVDTCFN 120
Db 61 PTMAOMKALISGFEETRCYGRIBGVVPIRHPNSICANNVGYITLSTSOVDTCFN 120
Qy 121 ASAPBEDCTSVTDLPNAPFDGPIITITVNRDGTTRYVQKGEYRNPEDIPSPNTDDVSS 180
Db 121 ASAPBEDCTSVTDLPNAPFDGPIITITVNRDGTTRYVQKGEYRNPEDIPSPNTDDVSS 180
Qy 181 GSSSERSTSGGYFFYFSTVHP1PDEDSPWITDSTRIPATNMDSSHSTTLOPTANPNT 240
Db 181 GSSSERSTSGGYFFYFSTVHP1PDEDSPWITDSTRIPATNMDSSHSTTLOPTANPNT 240
Qy 241 GLVEDLDRTGPIPLSMTTQOQNSQSPSTSHGLEDKDPHTTSTLTSSNRNDVTGRRDPNH 300
Db 241 GLVEDLDRTGPIPLSMTTQOQNSQSPSTSHGLEDKDPHTTSTLTSSNRNDVTGRRDPNH 300
Qy 301 SEGSTTLLEGYTSHPYTKESRTPIPTSAKTSFGVATVAGDSNNVNRSLSGDDPTF 360
Db 301 SEGSTTLLEGYTSHPYTKESRTPIPTSAKTSFGVATVAGDSNNVNRSLSGDDPTF 360
Qy 361 HPSGSHRTTHGSESDGSHSGSQEGGANTTSGPIRTPQIPBWLIIILASLLALAILAVCIA 420
Db 361 HPSGSHRTTHGSESDGSHSGSQEGGANTTSGPIRTPQIPBWLIIILASLLALAILAVCIA 420
Qy 421 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKQEMVHLVNRKESSTPDQFTADE 480
Db 421 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKQEMVHLVNRKESSTPDQFTADE 480
Qy 481 TRNLQNVDMKIGV 493
Db 481 TRNLQNVDMKIGV 493

RESULT 8
AAU02449
ID AAU02449 standard; protein; 493 AA.
AC AAU02449;

```

XX	29-AUG-2001 (first entry)	
DT		
XX	Human epithelial antigen CD44 polypeptide.	
DE		
XX	Human, epithelial antigen; immune-mediated disease; CD44; infection;	
KW	immune deficiency disorder; hypersensitivity; inflammation;	
KW	systemic lupus erythematosus; platelet disorder; rheumatoid arthritis;	
KW	transplant rejection; asthma.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	Domain	1..400
FT		/label= Extracellular_domain
FT	Modified-site	65..67
FT		/note= "N-linked glycosylation site"
FT	Modified-site	97..99
FT		/note= "N-linked glycosylation site"
FT	Modified-site	140..142
FT		/note= "N-linked glycosylation site"
FT	Modified-site	150..152
FT		/note= "N-linked glycosylation site"
FT	Modified-site	160..162
FT		/note= "N-linked glycosylation site"
FT	Binding-site	180..181
FT		/note= "SG dipeptide that forms the minimal attachment site of serine-linked chondroitin sulphate in proteoglycan proteins"
FT	Binding-site	190..191
FT		/note= "SG dipeptide that forms the minimal attachment site of serine-linked chondroitin sulphate in proteoglycan proteins"
FT	Binding-site	390..391
FT		/note= "SG dipeptide that forms the minimal attachment site of serine-linked chondroitin sulphate in proteoglycan proteins"
FT	Domain	401..421
FT		/label= Transmembrane_domain
FT	Modified-site	427..429
FT		/note= "N-linked glycosylation site"
PN	US6218525-B1.	
XX		
PD	17-APR-2001.	
XX		
PF	01-DEC-1992;	92US-00983647.
XX		
PR	25-FEB-1988;	88US-00160416.
PR	13-JUL-1989;	89US-00379076.
PR	13-JUL-1990;	90US-00553755.
XX		
PA	(GENO ) GEN HOSPITAL CORP.	
XX		
E1	Seed B, Arruffo A, Simmons D;	
XX		
XX	WPI; 2001-289848/30.	
DR	N-PSDB; AAS03192.	
XX		
PT	New recombinant DNA encoding CD28 useful for diagnosing and treating	
PT	immune-mediated diseases, infections or disorders, e.g. systemic lupus	
PT	erythematosus, asthma, transplant rejection, rheumatoid arthritis.	
XX		
PS	Example 15; Col 69-72; 72pp; English.	
CC	The present sequence encoding for human epithelial antigen CD44 is 1 of	
CC	various human lymphocyte cell surface antigen polypeptide sequences	
CC	(AAU02433-AAU02445) described in the present invention. The invention	
CC	relates to a novel method of cloning cDNA encoding cell surface antigen	
CC	and efficient construction of cDNA libraries. Also described are 2	
CC	expression vectors (AAS03171, AAS03174) which provide high level	
CC	expression in eukaryotic host cells. A genetically engineered cDNA	
CC	sequence encoding the CD28 amino acid extracellular domain sequence	

Query	Match	99.3%	Score 2590;	DB 4;	Length 493;
Best Local Similarity	99.4%	Pred. No. 2.9e-190;			
Matches 499;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0	
QY	1 MDKFTWHAHWGCLVPLISLAQIDINTCRFAGVYHVKNRYSISRTEADLCANSTL	60			
Db	1 MDKFTWHAHWGCLVPLISLAQIDINTCRFAGVYHVKNRYSISRTEADLCANSTL	60			
QY	61 PTMAQWEKALSIGFETCRVGFIEGHVYIPRIHPSICANNTGVYILTSNTSQDYTCFN	120			
Db	61 PTMAQWEKALSIGFETCRVGFIEGHVYIPRIHPSICANNTGVYILTSNTSQDYTCFN	120			
QY	121 ASAPREDCTSTYDLPNAPFDGPIITITTVNRDGRVYVKGGRYRNPEDYPSNPTDDVSS	180			
Db	121 ASAPREDCTSTYDLPNAPFDGPIITITTVNRDGRVYVKGGRYRNPEDYPSNPTDDVSS	180			
QY	181 GSSSESSSTGSGVIFYETFSYVHPIDPDESPWITDSTRIPATMDSHSTLLQPTANPT	240			
Db	181 GSSSESSSTGSGVIFYETFSYVHPIDPDESPWITDSTRIPATMDSHSTLLQPTANPT	240			
QY	241 GLVEDLDRGPILSMTTQOONSQSFSTSHGEGLEBKHPITTSITLSSNRDVTGGRDPNH	300			
Db	241 GLVEDLDRGPILSMTTQOONSQSFSTSHGEGLEBKHPITTSITLSSNRDVTGGRDPNH	300			
QY	301 SEGSTTLLEGYSHTPHTKESRTEPIVTSATGSGFVYATVVDSSNVNRSISGQDTF	360			
Db	301 SEGSTTLLEGYSHTPHTKESRTEPIVTSATGSGFVYATVVDSSNVNRSISGQDTF	360			
QY	361 HPGGSGHTHGESGSHGSHQOEGCANTTSGPITPPIPMWILLASLALAILIAYCTA	420			
Db	361 HPGGSGHTHGESGSHGSHQOEGCANTTSGPITPPIPMWILLASLALAILIAYCTA	420			
QY	421 VNSRRRCGKKCLVINSNGANGVEDRKPDSGLNGEASKSQEWVHLVNTSESTPPQFWTAD	480			
Db	421 VNSRRRCGKKCLVINSNGANGVEDRKPDSGLNGEASKSQEWVHLVNTSESTPPQFWTAD	480			
QY	481 TRTLQVMDKIGV	493			
Db	481 TRTLQVMDKIGV	493			
Sequence 493 AA;					
Query Match 99.3%; Score 2590; DB 4; Length 493;					
Best Local Similarity 99.4%; Pred. No. 2.9e-190;					
Matches 499; Conservative 0; Mismatches 3; Indels 0; Gaps 0					
RESULT 9					
ABU04637					
ID ABU04637 standard; protein; 493 AA.					
XX ABU04637;					
XX 29-JAN-2003 (first entry)					
XX Human expressed protein tag (BPT) #1303.					
XX Translational profiling; expressed protein tag; BPT; kinase; phosphatase;					
XX protease; protease inhibitor; transporter; cytoskeletal protein;					
XX receptor; transcription factor; cancer; MHC;					
XX major histocompatibility complex; myeloma; colon cancer; gastric cancer;					
XX adenocarcinoma; sarcoma; melanoma; lymphoma; leukemia.					
XX Homo sapiens.					

XX WO200278524-A2.  
 PN 10-OCT-2002.  
 XX 28-MAR-2002; 2002WO-US009671.  
 XX 28-MAR-2001; 2001US-0279495P.  
 PR 21-MAY-2001; 2001US-0292544P.  
 PR 08-AUG-2001; 2001US-0310801P.  
 PR 01-OCT-2001; 2001US-0326370P.  
 PR 04-DEC-2001; 2001US-0336780P.  
 PR 20-FEB-2002; 2002US-0358985P.  
 XX (ZYCO-) ZYCOS INC.  
 PA Chicx RM, Tomlinson AJ, Urban RG;  
 PI WPI; 2003-040607/03.  
 DR New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 XX cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.  
 XX Example 2; SEQ ID NO 1303; 134pp; English.  
 PS The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 493 AA;

Query Match 99.3%; Score 2590; DB 6; Length 493;  
 Best Local Similarity 99.4%; Pred. No. 2.9e-190;  
 Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MDRKFWHAAWGLCLVPLSLAQIDINTTCRPAGVHVHKNRGRYISRTAADLCKANSTL 60  
 DB 1 MDRKFWHAAWGLCLVPLSLAQIDINTTCRPAGVHVHKNRGRYISRTAADLCKANSTL 60  
 QY PTMAOMERKALSIGETGRYGPBEGHVIPIPHNSICANNNGVILTSWTSQYDYTCN 120  
 DB 61 PTMAOMERKALSIGETGRYGPBEGHVIPIPHNSICANNNGVILTSWTSQYDYTCN 120  
 QY 121 ASAPPEEDCTSVTDLPNAFDPIITITVNRDGRYVKGERTNPEDIPSNPTDDVSS 180  
 DB 121 ASAPPEEDCTSVTDLPNAFDPIITITVNRDGRYVKGERTNPEDIPSNPTDDVSS 180  
 QY 181 GSSSERSTSGGYFYFSTYTHPIEDDSPIWTSRIRATNMDSHSTTLQPTANPT 240  
 DB 181 GSSSERSTSGGYFYFSTYTHPIEDDSPIWTSRIRATNMDSHSTTLQPTANPT 240  
 QY 181 GSSSERSTSGGYFYFSTYTHPIEDDSPIWTSRIRATNMDSHSTTLQPTANPT 240  
 DB 241 GLVEDDLRTGSLMTTQSNQSFSTSHGLEDKXHPPTSTLSSNRNDVTGRRDPNH 300  
 QY 241 GLVEDDLRTGSLMTTQSNQSFSTSHGLEDKXHPPTSTLSSNRNDVTGRRDPNH 300  
 DB 301 SEGSTTLLEGYTHYPTKESRTIPVTSAKTSGFGVTAATVGDGNSNVNRSLSGQDPF 360  
 QY 301 SEGSTTLLEGYTHYPTKESRTIPVTSAKTSGFGVTAATVGDGNSNVNRSLSGQDPF 360

DB 301 SEGSTTLLEGYTHYPTKESRTIPVTSAKTSGFGVTAATVGDGNSNVNRSLSGQDPF 360  
 QY 361 HPSGSHRTTGSSESDGSHSQEGCANTTSGPIRTPQIPEMILLASLALAILAVCIA 420  
 DB 361 HPSGSHRTTGSSESDGSHSQEGCANTTSGPIRTPQIPEMILLASLALAILAVCIA 420  
 QY 421 VNSRRRCGCKKLVINGNANAVDRKPSGLNGASAKSQEWVHLWNKSSPTPCQFMTADE 480  
 DB 421 VNSRRRCGCKKLVINGNANAVDRKPSGLNGASAKSQEWVHLWNKSSPTPCQFMTADE 480  
 QY 481 TRNLQNDYDMKIGV 493  
 DB 481 TRNLQNDYDMKIGV 493  
 RESULT 10  
 ABU04627  
 ID ABU04627 standard; protein; 493 AA.  
 XX AC ABU04627;  
 XX 29-JAN-2003 (first entry)  
 DT Human expressed protein tag (EPT) #1293.  
 XX DE  
 XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
 KW protease; protease inhibitor; transporter; cytoskeletal protein;  
 KW receptor; transcription factor; cancer; MHC;  
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukemia.  
 XX OS Homo sapiens.  
 XX PN WO200278524-A2.  
 PD 10-OCT-2002.  
 XX PF 28-MAR-2002; 2002WO-US009671.  
 XX PR 28-MAR-2001; 2001US-0279495P.  
 PR 21-MAY-2001; 2001US-0292544P.  
 PR 08-AUG-2001; 2001US-0310801P.  
 PR 01-OCT-2001; 2001US-0326370P.  
 PR 04-DEC-2001; 2001US-0336780P.  
 PR 20-FEB-2002; 2002US-0358985P.  
 XX PA (ZYCO-) ZYCOS INC.  
 XX Chicx RM, Tomlinson AJ, Urban RG;  
 PI WPI; 2003-040607/03.  
 DR New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 XX cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.  
 XX Example 2; SEQ ID NO 1293; 134pp; English.  
 PS The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational

CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 493 AA;

Query Match 99.3%; Score 2590; DB 6; Length 493;

Best Local Similarity 99.4%; Pred. No. 2.9e-190;

Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKFWMAWAGLCIVPLSLAIDINITCRFAGVHYEKNGRYSISREADLCRAFNSTL 60  
 DB 1 MDKFWMAWAGLCIVPLSLAIDINITCRFAGVHYEKNGRYSISREADLCRAFNSTL 60  
 QY 61 PTMAQMEKALSIGETCRGYFIEGHVPIRHPNSICANNNGVYILTSNTSOYDTYCFN 120  
 DB 61 PTMAQMEKALSIGETCRGYFIEGHVPIRHPNSICANNNGVYILTSNTSOYDTYCFN 120  
 QY 121 ASAPPEDCTSVTDLPNAPDGPITITTVNRDGTTRYVQKGYRTNPEDIYPSNPTDDVSS 180  
 DB 121 ASAPPEDCTSVTDLPNAPDGPITITTVNRDGTTRYVQKGYRTNPEDIYPSNPTDDVSS 180  
 QY 181 GSSSERSSTSGGYFYFTSTVHPIDPDESPWITDSTRIPATNDSHSTTLQPTANPMT 240  
 DB 181 GSSSERSSTSGGYFYFTSTVHPIDPDESPWITDSTRIPATNDSHSTTLQPTANPMT 240  
 QY 241 GLVEDDLRTGTPLSMTTQOSNSQSFSTSHGLEDKDHPTTSTLTSSNRNDVYGRDPMH 300  
 DB 241 GLVEDDLRTGTPLSMTTQOSNSQSFSTSHGLEDKDHPTTSTLTSSNRNDVYGRDPMH 300  
 QY 301 SEGSTTLLEGYTSHPPTKESRTFIPYTSAKTSGFGLTAVTVDSNSVNRSLSGDDTF 360  
 DB 301 SEGSTTLLEGYTSHPPTKESRTFIPYTSAKTSGFGLTAVTVDSNSVNRSLSGDDTF 360  
 QY 361 HPSGSGSHTTGSESDGSHSGSQEGGANTTSGPIRTPOIPFWLIIILASLALAILAVCIA 420  
 DB 361 HPSGSGSHTTGSESDGSHSGSQEGGANTTSGPIRTPOIPFWLIIILASLALAILAVCIA 420  
 QY 421 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASQSEVHLVNRKSESTPDQFMTADE 480  
 DB 421 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASQSEVHLVNRKSESTPDQFMTADE 480  
 QY 481 TRNLQNVDMKIGV 493  
 DB 481 TRNLQNVDMKIGV 493

RESULT 11  
 ID ABU04639 standard; protein; 493 AA.  
 AC ABU04639;  
 XX 29-JAN-2003 (first entry)  
 DB Human expressed protein tag (EPT) #1305.  
 XX Translational profiling: expressed protein tag; EPT; kinase; phosphatase;  
 KW protease; protease inhibitor; transporter; cytoskeletal protein;  
 KW receptor; transcription factor; cancer; MHC;  
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO200278524-A2.  
 PD 10-OCT-2002.  
 XX  
 PF 28-MAR-2002; 2002MO-US009671.  
 XX 28-MAR-2001; 2001US-0279495P.  
 PR 21-MAY-2001; 2001US-0292544P.  
 PR

PR 08-AUG-2001; 2001US-0310801P.  
 PR 01-OCT-2001; 2001US-0326370P.  
 PR 04-DEC-2001; 2001US-0336780P.  
 PR 20-FEB-2002; 2002US-0358985P.

(ZYCO-) ZYCO INC.

Chicz RM, Tomlinson AJ, Urban RG;

WPI, 2003-040607/03.

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.

Example 2; SEQ ID NO 1305; 134p; English.

CC The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 493 AA;

Query Match 99.3%; Score 2590; DB 6; Length 493;

Best Local Similarity 99.4%; Pred. No. 2.9e-190;

Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKFWMAWAGLCIVPLSLAIDINITCRFAGVHYEKNGRYSISREADLCRAFNSTL 60  
 DB 1 MDKFWMAWAGLCIVPLSLAIDINITCRFAGVHYEKNGRYSISREADLCRAFNSTL 60  
 QY 61 PTMAQMEKALSIGETCRGYFIEGHVPIRHPNSICANNNGVYILTSNTSOYDTYCFN 120  
 DB 61 PTMAQMEKALSIGETCRGYFIEGHVPIRHPNSICANNNGVYILTSNTSOYDTYCFN 120  
 QY 121 ASAPPEDCTSVTDLPNAPDGPITITTVNRDGTTRYVQKGYRTNPEDIYPSNPTDDVSS 180  
 DB 121 ASAPPEDCTSVTDLPNAPDGPITITTVNRDGTTRYVQKGYRTNPEDIYPSNPTDDVSS 180  
 QY 181 GSSSERSSTSGGYFYFTSTVHPIDPDESPWITDSTRIPATNDSHSTTLQPTANPMT 240  
 DB 181 GSSSERSSTSGGYFYFTSTVHPIDPDESPWITDSTRIPATNDSHSTTLQPTANPMT 240  
 QY 241 GLVEDDLRTGTPLSMTTQOSNSQSFSTSHGLEDKDHPTTSTLTSSNRNDVYGRDPMH 300  
 DB 241 GLVEDDLRTGTPLSMTTQOSNSQSFSTSHGLEDKDHPTTSTLTSSNRNDVYGRDPMH 300  
 QY 301 SEGSTTLLEGYTSHPPTKESRTFIPYTSAKTSGFGLTAVTVDSNSVNRSLSGDDTF 360  
 DB 301 SEGSTTLLEGYTSHPPTKESRTFIPYTSAKTSGFGLTAVTVDSNSVNRSLSGDDTF 360  
 QY 361 HPSGSGSHTTGSESDGSHSGSQEGGANTTSGPIRTPOIPFWLIIILASLALAILAVCIA 420  
 DB 361 HPSGSGSHTTGSESDGSHSGSQEGGANTTSGPIRTPOIPFWLIIILASLALAILAVCIA 420  
 QY 421 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASQSEVHLVNRKSESTPDQFMTADE 480  
 DB 421 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASQSEVHLVNRKSESTPDQFMTADE 480

QY 481 TRNLQNVDMKIGV 493  
 DB 481 TRNLQNVDMKIGV 493

# RESULT 12

ABU04633 standard; protein; 493 AA.

AC ABU04633;

DT 29-JAN-2003 (first entry)

DE Human expressed protein tag (EPT) #1299.

KM Translational profiling; expressed protein tag; EPT; kinase; phosphatase;

KM protease; protease inhibitor; transporter; cytoskeletal protein;

KM receptor; transcription factor; cancer; MHC;

KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;

KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

OS Homo sapiens.

PN WO200278524-A2.

PD 10-OCT-2002.

PF 28-MAR-2002; 2002MO-US009671.

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

PA (ZYCO-) ZYCO INC.

PI Chiciz RM, Tomlinson AJ, Urban RG;

DR WPI; 2003-040607/03.

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.

PS Example 2, SEQ ID NO 1299; 134pp; English.

CC The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide binds also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukaemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 493 AA;

Query Match 99.3%; Score 2590; DB 6; Length 493;  
 Best Local Similarity 99.4%; Pred. No. 2, 9e-190;  
 Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKFWMAAMGLCLVPLSLAQIDINITCRFAGVHVENKGRYSISRTAADLCAFNSTL 60  
 DB 1 MDKFWMAAMGLCLVPLSLAQIDINITCRFAGVHVENKGRYSISRTAADLCAFNSTL 60  
 QY 61 PTMAQMEKALISGFETCYGVFIEGHVYIPRIHPNSICANNNTGYIILTSNSQYDYCFN 120  
 DB 61 PTMAQMEKALISGFETCYGVFIEGHVYIPRIHPNSICANNNTGYIILTSNSQYDYCFN 120  
 QY 121 ASAPPEEDCTSVTLDPNAPDGPITITIVNRDGTIRYVQXGEYRTNPEDITYENPTDDVSS 180  
 DB 121 ASAPPEEDCTSVTLDPNAPDGPITITIVNRDGTIRYVQXGEYRTNPEDITYENPTDDVSS 180  
 QY 181 GSSSERSSSTGGYIPTYESTVYHPIPDDESPWITSTDRIPATNDSHSTLQPTANPT 240  
 DB 181 GSSSERSSSTGGYIPTYESTVYHPIPDDESPWITSTDRIPATNDSHSTLQPTANPT 240  
 QY 241 GLVEDLDRTGPLSMVTQGSNSQSFSTSHGLLEDDEPTSTLTSSNNDVYTGRRDPNH 300  
 DB 241 GLVEDLDRTGPLSMVTQGSNSQSFSTSHGLLEDDEPTSTLTSSNNDVYTGRRDPNH 300  
 QY 301 SEGSTTLLEGYSHYPHTKERTPIPTSAKTGSGVTAIVGDSNVNSSLGDDQTF 360  
 DB 301 SEGSTTLLEGYSHYPHTKERTPIPTSAKTGSGVTAIVGDSNVNSSLGDDQTF 360  
 QY 361 HPGGSHHTHGSESDGSHSGOEGGANTSGIRTPQIPFWLIIIASLALALIAVCIA 420  
 DB 361 HPGGSHHTHGSESDGSHSGOEGGANTSGIRTPQIPFWLIIIASLALALIAVCIA 420  
 QY 421 VNSRRRCQCKKLIVNSGNGAVEDKRPKNGINGEASKSOEMHIVNKKSEETPDQMTDE 480  
 DB 421 VNSRRRCQCKKLIVNSGNGAVEDKRPKNGINGEASKSOEMHIVNKKSEETPDQMTDE 480  
 QY 481 TRNLQNVDMKIGV 493  
 DB 481 TRNLQNVDMKIGV 493

## RESULT 13

ABU04635 standard; protein; 493 AA.

AC ABU04635;

DT 29-JAN-2003 (first entry)

DE Human expressed protein tag (EPT) #1301.

KM Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
 KM protease; protease inhibitor; transporter; cytoskeletal protein;  
 KM receptor; transcription factor; cancer; MHC;  
 KM major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KM adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

OS Homo sapiens.

PN WO200278524-A2.

PD 10-OCT-2002.

PF 28-MAR-2002; 2002MO-US009671.

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

PA (ZYCO-) ZYCO INC.

PI Chiciz RM, Tomlinson AJ, Urban RG;

DR WPI, 2003-040607/03.  
 XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 XX leukemia.  
 XX  
 PS Example 2; SEQ ID NO 1301; 134pp; English.  
 XX  
 CC The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC lymphoma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC myeloma or leukemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 493 AA:  
 Query Match 99.3%; Score 2590; DB 6; Length 493;  
 Best Local Similarity 99.4%; Pred. No. 2.9e-190;  
 Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MDKFWHAAWGLCLVPLSLAQIDLNTICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60  
 DB 1 MDKFWHAAWGLCLVPLSLAQIDLNTICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60  
 QY 61 PTMAQMEKALSIGETCRGYGFIEGHVAVIPRIHNSICANNNGVYLLTNTSOYDYTCFN 120  
 DB 61 PTMAQMEKALSIGETCRGYGFIEGHVAVIPRIHNSICANNNGVYLLTNTSOYDYTCFN 120  
 QY 61 PTMAQMEKALSIGETCRGYGFIEGHVAVIPRIHNSICANNNGVYLLTNTSOYDYTCFN 120  
 DB 61 PTMAQMEKALSIGETCRGYGFIEGHVAVIPRIHNSICANNNGVYLLTNTSOYDYTCFN 120  
 QY 121 ASAPBEDCTSVTLDPNAPDGPITITIVNRDGTTRYOKGEYRTNPEIDYPSNPTDDVSS 180  
 DB 121 ASAPBEDCTSVTLDPNAPDGPITITIVNRDGTTRYOKGEYRTNPEIDYPSNPTDDVSS 180  
 QY 181 GSSSERSTSGGYIFTFSTVHPIDPDSFWITDSTRIPATMDSSHTTIOPTANPMT 240  
 DB 181 GSSSERSTSGGYIFTFSTVHPIDPDSFWITDSTRIPATMDSSHTTIOPTANPMT 240  
 QY 241 GVEDDLPRTPLSNTTQOSNSQSFSTSHGELKEDKHPTSTLTSNRNDVYGGRRDPNH 300  
 DB 241 GVEDDLPRTPLSNTTQOSNSQSFSTSHGELKEDKHPTSTLTSNRNDVYGGRRDPNH 300  
 QY 301 SEGSTTLLEGYTHYPTKSRPIFYTSAXTGSFGYAVTVGDSNNVRSISGQDPTF 360  
 DB 301 SEGSTTLLEGYTHYPTKSRPIFYTSAXTGSFGYAVTVGDSNNVRSISGQDPTF 360  
 QY 361 HPSGGSHHTGSESDGSHSGOGGANTTSGPRTQIPQIPLWLLIISLALMLTILAVCIA 420  
 DB 361 HPSGGSHHTGSESDGSHSGOGGANTTSGPRTQIPQIPLWLLIISLALMLTILAVCIA 420  
 QY 421 VNSRRRCQKKKLVINSNGAVDRKPSGLNGEASRQEMVHLVNKSESPTDQFWTAD 480  
 DB 421 VNSRRRCQKKKLVINSNGAVDRKPSGLNGEASRQEMVHLVNKSESPTDQFWTAD 480  
 QY 481 TRNLQNDMKIGV 493  
 DB 481 TRNLQNDMKIGV 493  
 DB 481 TRNLQNDMKIGV 493

XX  
 AC ABU04613;  
 XX  
 DT 29-JAN-2003 (first entry)  
 XX  
 DE Human expressed protein tag (EPT) #1279.  
 XX  
 KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
 KW protease; protease inhibitor; transporter; cytoskeletal protein;  
 KW receptor; transcription factor; cancer; MHC;  
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukemia.  
 XX  
 XX Homo sapiens.  
 XX  
 PD WO200278524-A2.  
 XX  
 XX 10-OCT-2002.  
 XX  
 XX 28-MAR-2002; 2002WO-US009671.  
 XX  
 XX 28-MAR-2001; 2001US-0279495P.  
 PR 21-MAY-2001; 2001US-0292544P.  
 PR 08-AUG-2001; 2001US-0310801P.  
 PR 01-OCT-2001; 2001US-0326370P.  
 PR 04-DEC-2001; 2001US-0356780P.  
 PR 20-FEB-2002; 2002US-0358985P.  
 XX  
 XX (ZYCO-) ZYCOs INC.  
 XX  
 XX Chicx RM, Tomlinson AJ, Urban RG;  
 XX  
 DR WPI, 2003-040607/03.  
 XX  
 PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.  
 XX  
 PS Example 2; SEQ ID NO 1279; 134pp; English.  
 XX  
 CC The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC lymphoma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC myeloma or leukemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 493 AA:  
 Query Match 99.3%; Score 2590; DB 6; Length 493;  
 Best Local Similarity 99.4%; Pred. No. 2.9e-190;  
 Matches 490; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MDKFWHAAWGLCLVPLSLAQIDLNTICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60  
 DB 1 MDKFWHAAWGLCLVPLSLAQIDLNTICRFAGVHVEKNGRYSISRTAADLCKAFNSTL 60  
 QY 61 PTMAQMEKALSIGETCRGYGFIEGHVAVIPRIHNSICANNNGVYLLTNTSOYDYTCFN 120  
 DB 61 PTMAQMEKALSIGETCRGYGFIEGHVAVIPRIHNSICANNNGVYLLTNTSOYDYTCFN 120





Tue Mar 9 08:08:40 2004

pct-us01-51014-1.rag

Page 13

QY	481	TRNLQNVDMKIGV	493
Db	481	TRNLQNVDMKIGV	493

Search completed: March 8, 2004, 06:10:03  
Job time : 62 secs

Tue Mar 9 08:08:41 2004

pct-us01-51014-1.rapb

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2004, 06:10:05 ; Search time 397 Seconds  
(without alignments)  
262.213 Million cell updates/sec

Title: PCT-US01-51014-1

Perfect score: 1 MDKFWHAAAGCLVPLSLA.....QMTADETRLQVDMKIV 493

Sequence:

Scoring table: BLOSUM62

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications\_AA.\*

1: /cgn2\_6/prodata/2/pubppa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/prodata/2/pubppa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/2/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/2/pubppa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/2/pubppa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/2/pubppa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/2/pubppa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/2/pubppa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/2/pubppa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/prodata/2/pubppa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/2/pubppa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/prodata/2/pubppa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/prodata/2/pubppa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/prodata/2/pubppa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/prodata/2/pubppa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/prodata/2/pubppa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/prodata/2/pubppa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/prodata/2/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2489.5	95.5	700	14	US-10-012-969C-2
2	2468.5	94.7	742	10	US-09-983-000A-16
3	2462.5	94.4	742	15	US-10-116-275-206
4	1361	52.2	365	9	US-09-870-759-73
5	1361	52.2	365	10	US-09-751-708A-73
6	1299	48.8	339	14	US-10-291-634-3
7	1299	48.8	339	14	US-10-156-932-18
8	734.5	28.5	143	14	US-10-156-932-20
9	718	27.5	185	14	US-10-156-932-22
10	537	20.6	109	9	US-10-156-932-16
11	478	18.3	90	4	US-09-799-118-3
12	473	18.1	90	10	US-09-927-463-8
13	386	14.8	79	14	US-10-156-932-14
14	349	13.4	67	14	US-10-052-641-2
15	199.5	7.6	322	9	US-09-909-320-201

16	199.5	7.6	322	9	US-09-909-088B-201	Sequence 201, App
17	199.5	7.6	322	9	US-09-905-291A-201	Sequence 201, App
18	199.5	7.6	322	9	US-09-902-853-201	Sequence 201, App
19	199.5	7.6	322	9	US-09-907-824-201	Sequence 201, App
20	199.5	7.6	322	9	US-09-907-841-201	Sequence 201, App
21	199.5	7.6	322	10	US-09-904-011-201	Sequence 201, App
22	199.5	7.6	322	10	US-09-906-838-201	Sequence 201, App
23	199.5	7.6	322	10	US-09-907-613-201	Sequence 201, App
24	199.5	7.6	322	10	US-09-904-942-201	Sequence 201, App
25	199.5	7.6	322	10	US-09-904-859-201	Sequence 201, App
26	199.5	7.6	322	10	US-09-904-820-201	Sequence 201, App
27	199.5	7.6	322	10	US-09-904-786-201	Sequence 201, App
28	199.5	7.6	322	10	US-09-904-736-201	Sequence 201, App
29	199.5	7.6	322	10	US-09-906-700-201	Sequence 201, App
30	199.5	7.6	322	10	US-09-903-786-201	Sequence 201, App
31	199.5	7.6	322	10	US-09-902-903-201	Sequence 201, App
32	199.5	7.6	322	10	US-09-903-749A-201	Sequence 201, App
33	199.5	7.6	322	10	US-09-904-462-201	Sequence 201, App
34	199.5	7.6	322	10	US-09-907-925-201	Sequence 201, App
35	199.5	7.6	322	10	US-09-902-692-201	Sequence 201, App
36	199.5	7.6	322	10	US-09-903-520-201	Sequence 201, App
37	199.5	7.6	322	10	US-09-905-056-201	Sequence 201, App
38	199.5	7.6	322	10	US-09-907-794-201	Sequence 201, App
39	199.5	7.6	322	10	US-09-903-943-201	Sequence 201, App
40	199.5	7.6	322	10	US-09-904-462-201	Sequence 201, App
41	199.5	7.6	322	10	US-09-907-925-201	Sequence 201, App
42	199.5	7.6	322	10	US-09-902-692-201	Sequence 201, App
43	199.5	7.6	322	10	US-09-903-520-201	Sequence 201, App
44	199.5	7.6	322	10	US-09-905-056-201	Sequence 201, App
45	199.5	7.6	322	10	US-09-909-064-201	Sequence 201, App

#### ALIGNMENTS

RESULT 1  
US-10-012-969C-2  
; Sequence 2, Application US/10012969C  
; Publication No. US20030108984A1  
; GENERAL INFORMATION:  
; APPLICANT: NAOR, DAVID  
; APPLICANT: GOLAN, ITSHAK  
; APPLICANT: NEVEVETKI, SHLOMO  
; TITLE OF INVENTION: C944 SPLICE VARIANT ASSOCIATED WITH RHEUMATOID ARTHRITIS  
; FILE REFERENCE: C00L21-2  
; CURRENT FILING DATE: 2002-09-25  
; PRIOR APPLICATION NUMBER: PCT/IL00/00326  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 700  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-012-969C-2

Query Match 95.5%; Score 2489.5; DB 14; Length 700;

Best Local Similarity 70.3%; Pred. No. 6.5e-190; Indels 207; Gaps 1;  
Matches 492; Conservative 0;  
1 MDKFWHAAAGCLVPLSLAIDNITCRFAGVHVEKNGRYSISRTADLCANFSTL 60  
1 MDKFWHAAAGCLVPLSLAIDNITCRFAGVHVEKNGRYSISRTADLCANFSTL 60  
1 PTMAOMERALSIGFTCYGFIKGVVIFRIHNSICANNITGVYILTSNTSXYDTYCFN 120  
61 PTMAOMERALSIGFTCYGFIKGVVIFRIHNSICANNITGVYILTSNTSXYDTYCFN 120  
121 ASAPDEEDCTSVTLDFNAPDPIITITIVNRDGTIRYVKGERTNPEDYIPSNPTDDVSS 180  
121 ASAPDEEDCTSVTLDFNAPDPIITITIVNRDGTIRYVKGERTNPEDYIPSNPTDDVSS 180

QY	181	GSSSESSSSGAYIVETSTVPIEDDSPIITSTDRPAT	222
Db	181	GSSSESSSTSGGIIFTTSTVHPIDEDSPWITSDRIIPATSTSSNTISAGMEPNEENE	240
QY	223	-----	222
Db	241	DERDRLLSPSGCIDDEDDFISSTISTTPRAFDHTKQNDWTQWNPNSHNPVLLQTTTR	3000
QY	223	-----	222
Db	301	MTADVDRNGTTAYEGNMNPBAHPPLIHHEHHEEETPHSTSTIQATPSSTTEBTATQEQ	360
QY	223	-----	222
Db	361	WPNRMHBEGRQTPREDSSHITGTAAASAHTSHPMQGRTPSPEDSSWTDFNFI SHPMG	420
QY	223	-----	222
Db	421	RGHQAQRMDMSHSTTLQPTANPMTGIVELDITGPILSMTQOONSQSSSTISHBGLBE	480
QY	274	DKDHPITSTLSSNRDVTGGRDPNHSBGSITLLEGYTSHYPHTKSRTEIPVTSATKG	333
Db	481	DKDHPITSTLSSNRDVTGGRDPNHSBGSITLLEGYTSHYPHTKSRTEIPVTSATKG	540
QY	334	SPGVTAVTVGDSNNRSLSGDQDPFHSSGSHHTHSESDGSHSGQEGAGWTSGP	393
Db	541	SPGVTAVTVGDSNNRSLSGDQDPFHSSGSHHTHSESDGSHSGQEGAGWTSGP	600
QY	394	RTEQIPBWLIIILASLALALIIAVCIANVSRRCCQKKLVINSNGAVEDRKPSGLNGE	453
Db	601	RTEQIPBWLIIILASLALALIIAVCIANVSRRCCQKKLVINSNGAVEDRKPSGLNGE	660
QY	454	ASKSQEWVHLVYNKESSETPDQPTADDETNLQNDMKICGV	493
Db	661	ASKSQEWVHLVYNKESSETPDQPTADDETNLQNDMKICGV	700

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

NAME/KEY: DOMAIN
LOCATION: (671)..(742)
OTHER INFORMATION: CYTOPLASMIC (POTENTIAL)
NAME/KEY: DOMAIN
LOCATION: (48)..(119)
OTHER INFORMATION: LINK
NAME/KEY: MOD_RBS
LOCATION: (21)..(21)
OTHER INFORMATION: PYROLIDONE CARBOXYLIC ACID (PROBABLE)
NAME/KEY: DISULFID
LOCATION: (53)..(118)
OTHER INFORMATION: BY SIMILARITY
NAME/KEY: DISULFID
LOCATION: (77)..(97)
OTHER INFORMATION: BY SIMILARITY
NAME/KEY: CARBOHYD
LOCATION: (25)..(25)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (57)..(57)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (100)..(100)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (110)..(110)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (120)..(120)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (350)..(350)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (548)..(548)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (599)..(599)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: CARBOHYD
LOCATION: (636)..(636)
OTHER INFORMATION: N-LINKED (GLCNAC ...) (POTENTIAL)
NAME/KEY: VARSPPLIC
LOCATION: (23)..(29)
OTHER INFORMATION: DNITCR -> GVGRRKS (IN ISOFORM CD44SP)
NAME/KEY: VARSPPLIC
LOCATION: (30)..(742)
OTHER INFORMATION: MISSING (IN ISOFORM CD44SP)
NAME/KEY: VARSPPLIC
LOCATION: (192)..(192)
OTHER INFORMATION: G -> A (IN ISOFORM WITH ALTERNATIVE SPLIC DONOR/ACCEPTOR ON)
NAME/KEY: VARSPPLIC
LOCATION: (5)
OTHER INFORMATION: 5
NAME/KEY: VARSPPLIC
LOCATION: (193)..(223)
OTHER INFORMATION: MISSING (IN ISOFORM WITH ALTERNATIVE SPLIC DONOR/ACCEPTOR ON)
NAME/KEY: VARSPPLIC
LOCATION: (223)..(223)
OTHER INFORMATION: T -> S (IN ISOFORM WITHOUT EXON 6)
NAME/KEY: VARSPPLIC
LOCATION: (224)..(266)
OTHER INFORMATION: MISSING (IN ISOFORM WITHOUT EXON 6)
NAME/KEY: VARSPPLIC
LOCATION: (223)..(223)
OTHER INFORMATION: T -> N (IN ISOFORMS WITHOUT EXONS 6 TO 11)
NAME/KEY: VARSPPLIC
LOCATION: (224)..(472)
OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXONS 6 TO 11)
NAME/KEY: VARSPPLIC
LOCATION: (223)..(223)
OTHER INFORMATION: T -> R (IN ISOFORMS WITHOUT EXONS 6 TO 14)
NAME/KEY: VARSPPLIC
LOCATION: (224)..(604)

```

```

OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXONS 6 TO 14)
NAME/KEY: VARSPLIC
LOCATION: (266) ..(273)
OTHER INFORMATION: MISSING (IN ISOFORM WITH ALTERNATIVE SPLIC DONOR/ACCEPTOR ON EXC
OTHER INFORMATION: N 7)
NAME/KEY: VARSPLIC
LOCATION: (385) ..(385)
OTHER INFORMATION: I -> T (IN ISOFORMS WITHOUT EXON 10)
NAME/KEY: VARSPLIC
LOCATION: (386) ..(428)
OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXON 10)
NAME/KEY: VARSPLIC
LOCATION: (506) ..(506)
OTHER INFORMATION: Q -> R (IN ISOFORMS WITHOUT EXON 13)
NAME/KEY: VARSPLIC
LOCATION: (507) ..(535)
OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXON 13)
NAME/KEY: VARSPLIC
LOCATION: (536) ..(536)
OTHER INFORMATION: N -> R (IN ISOFORMS WITHOUT EXON 14)
NAME/KEY: VARSPLIC
LOCATION: (537) ..(604)
OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXON 14)
NAME/KEY: VARSPLIC
LOCATION: (675) ..(675)
OTHER INFORMATION: R -> S (IN ISOFORMS WITHOUT EXON 19)
NAME/KEY: VARSPLIC
LOCATION: (676) ..(742)
OTHER INFORMATION: MISSING (IN ISOFORMS WITHOUT EXON 19)
NAME/KEY: VARIANT
LOCATION: (46) ..(46)
OTHER INFORMATION: R -> P (IN IN(A) ANTIGEN)
US-09-983-000A-16

```

Query Match 94.7%; Score 2468.5; DB 10; Length 742;  
 Best Local Similarity 66.3%; Pred. No. 3.3e-188;  
 Matches 492; Conservative 1; Mismatches 0; Indels 249; Gaps 1;

```

1 MDKFWMAAMGLCLVPLSLAQIDLNITCRFAGVFNHVKNGRYSISRTAADLCKAFNSTL 60
1 MDKFWMAAMGLCLVPLSLAQIDLNITCRFAGVFNHVKNGRYSISRTAADLCKAFNSTL 60
61 PTMAQMEKALISIGETCRYGFIHGHVVIPIHNSICAANNVGYIILNSTQYDTYCFN 120
61 PTMAQMEKALISIGETCRYGFIHGHVVIPIHNSICAANNVGYIILNSTQYDTYCFN 120
121 ASAPPEEDCTSVTLDPNAPDGPITITIVNDDGTRVYQKGEYRNPEDIYSPNPTDDVSS 180
121 ASAPPEEDCTSVTLDPNAPDGPITITIVNDDGTRVYQKGEYRNPEDIYSPNPTDDVSS 180
181 GSSSERSTSGGYIFTFSTVHPIDPDSFWITDSTRIPAT----- 222
181 GSSSERSTSGGYIFTFSTVHPIDPDSFWITDSTRIPAT----- 222
223 ----- 222
223 ----- 222
241 TWDWFSWFLPSESXKHLHTTQMACTSSNTISAGWEPNENEDERDRHLSFGSGGIDDD 300
241 TWDWFSWFLPSESXKHLHTTQMACTSSNTISAGWEPNENEDERDRHLSFGSGGIDDD 300
223 ----- 222
301 EDFISSTISTPRAFDHTKONQDWTQWNPESHNEVLLQTTMTVDVDRNGTAYEGNMN 360
301 EDFISSTISTPRAFDHTKONQDWTQWNPESHNEVLLQTTMTVDVDRNGTAYEGNMN 360
223 ----- 222
421 HSTTGTAASAHSHPMQGRTPSPEDSSWTFENPISHPMGRGHQAGRRMDSDSHSTT 480
421 HSTTGTAASAHSHPMQGRTPSPEDSSWTFENPISHPMGRGHQAGRRMDSDSHSTT 480
223 -----NMDSHSTT 231
232 LQTPANNTGVLVEDLDKTPGLSMTOOSNOSFSTSHGLLEDKHPTTSTLLTSRRNV 291
232 LQTPANNTGVLVEDLDKTPGLSMTOOSNOSFSTSHGLLEDKHPTTSTLLTSRRNV 291
481 LQTPANNTGVLVEDLDKTPGLSMTOOSNOSFSTSHGLLEDKHPTTSTLLTSRRNV 540
481 LQTPANNTGVLVEDLDKTPGLSMTOOSNOSFSTSHGLLEDKHPTTSTLLTSRRNV 540

```

```

292 TGRGRDNHSGSTTLLEGYTSHPATKESRTFIPYSAKTSGFGVTAIVTGDSNSVNR 351
541 TGRGRDNHSGSTTLLEGYTSHPATKESRTFIPYSAKTSGFGVTAIVTGDSNSVNR 600
352 SLSDGDTFHPGSGSHHTHSGSDGSHSGOEGANTSGPIRTPQPELIIASLAL 411
601 SLSDGDTFHPGSGSHHTHSGSDGSHSGOEGANTSGPIRTPQPELIIASLAL 660
412 ALIIVACIAVNSRRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVVKESSET 471
661 ALIIVACIAVNSRRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVVKESSET 720
472 PDQFMTADETRNLQNDMKIGV 493
721 PDQFMTADETRNLQNDMKIGV 742

```

## RESULT 3

```

US-10-116-275-206
Sequence 206, Application US/10116275
Publication No. US2003021476A1
GENERAL INFORMATION:
APPLICANT: Elian Pharmaceutical Technology
APPLICANT: O'Mahony, Daniel J.
APPLICANT: Brayden, David
APPLICANT: Byrne, Daragh
APPLICANT: Lambkin, Imelda
APPLICANT: Higgins, Lisa
TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and Methods
FILE REFERENCE: E1067/20087
CURRENT APPLICATION NUMBER: US/10/116,275
CURRENT FILING DATE: 2002-10-04
NUMBER OF SEQ ID NOS: 349
SOFTWARE: PatentIn version 3.1
SEQ ID NO 206
LENGTH: 742
TYPE: PRT
ORGANISM: Homo sapiens
US-10-116-275-206

```

Query Match 94.4%; Score 2462.5; DB 15; Length 742;  
 Best Local Similarity 66.2%; Pred. No. 1e-187;  
 Matches 491; Conservative 1; Mismatches 1; Indels 249; Gaps 1;

```

1 MDKFWMAAMGLCLVPLSLAQIDLNITCRFAGVFNHVKNGRYSISRTAADLCKAFNSTL 60
1 MDKFWMAAMGLCLVPLSLAQIDLNITCRFAGVFNHVKNGRYSISRTAADLCKAFNSTL 60
61 PTMAQMEKALISIGETCRYGFIHGHVVIPIHNSICAANNVGYIILNSTQYDTYCFN 120
61 PTMAQMEKALISIGETCRYGFIHGHVVIPIHNSICAANNVGYIILNSTQYDTYCFN 120
121 ASAPPEEDCTSVTLDPNAPDGPITITIVNDDGTRVYQKGEYRNPEDIYSPNPTDDVSS 180
121 ASAPPEEDCTSVTLDPNAPDGPITITIVNDDGTRVYQKGEYRNPEDIYSPNPTDDVSS 180
181 GSSSERSTSGGYIFTFSTVHPIDPDSFWITDSTRIPAT----- 222
181 GSSSERSTSGGYIFTFSTVHPIDPDSFWITDSTRIPAT----- 222
223 ----- 222
223 ----- 222
241 TWDWFSWFLPSESXKHLHTTQMACTSSNTISAGWEPNENEDERDRHLSFGSGGIDDD 300
241 TWDWFSWFLPSESXKHLHTTQMACTSSNTISAGWEPNENEDERDRHLSFGSGGIDDD 300
223 ----- 222
301 EDFISSTISTPRAFDHTKONQDWTQWNPESHNEVLLQTTMTVDVDRNGTAYEGNMN 360
301 EDFISSTISTPRAFDHTKONQDWTQWNPESHNEVLLQTTMTVDVDRNGTAYEGNMN 360
223 ----- 222
421 HSTTGTAASAHSHPMQGRTPSPEDSSWTFENPISHPMGRGHQAGRRMDSDSHSTT 480
421 HSTTGTAASAHSHPMQGRTPSPEDSSWTFENPISHPMGRGHQAGRRMDSDSHSTT 480
223 ----- 222
481 LQTPANNTGVLVEDLDKTPGLSMTOOSNOSFSTSHGLLEDKHPTTSTLLTSRRNV 540
481 LQTPANNTGVLVEDLDKTPGLSMTOOSNOSFSTSHGLLEDKHPTTSTLLTSRRNV 540

```

QY 223 -----NMDSHSTT 231  
DB 421 HSTGTAAASHTSHPMQGRRTTSPEDSWTDFNPI SHPMGRHQAGRMMDSSHT 480  
QY 232 LQFANPNTGLVDELDTGELSMTTOQSNOSQSTSHGEGLEEDKHPHTSTLSSNRNV 291  
DB 481 LQFANPNTGLVDELDTGELSMTTOQSNOSQSTSHGEGLEEDKHPHTSTLSSNRNV 540  
QY 292 TGRARDPNHSGSTTLLEGYTSYHTKESRTFIPYTSAKTGSFGYAVTVGDSNVNR 351  
DB 541 TGRARDPNHSGSTTLLEGYTSYHTKESRTFIPYTSAKTGSFGYAVTVGDSNVNR 600  
QY 352 SLSDODTFHPSGSGSHTHGSESDHSGSQEGGANTTSGPIPTPOIPEWLLILASLAL 411  
DB 601 SLSDODTFHPSGSGSHTHGSESDHSGSQEGGANTTSGPIPTPOIPEWLLILASLAL 660  
QY 412 ALLAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNSESST 471  
DB 661 ALLAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNSESST 720  
QY 472 PDQFMTADETRNLOVDMKIGV 493  
DB 721 PDQFMTADETRNLOVDMKIGV 742

## RESULT 4

US-09-870-759-73  
; Sequence 73, Application US/09870759  
; Patent No. US2002017551A1  
; GENERAL INFORMATION:  
; APPLICANT: TERMAN, David S  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
; FILE REFERENCE: 870759  
; CURRENT APPLICATION NUMBER: US/09/870,759  
; CURRENT FILING DATE: 2002-01-14  
; PRIOR APPLICATION NUMBER: US 60/208,128  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 166  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 73  
; LENGTH: 365  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-870-759-73

Query Match 52.2%; Score 1361; DB 9; Length 365;  
Best Local Similarity 57.7%; Pred. No. 2.8e-100;  
Matches 288; Conservative 28; Mismatches 43; Indels 140; Gaps 9;

QY 1 MDKFWMTAAWGLCLVPLSLA-----QIDLNTTCRFAGVFHYEKNRGYSISRTAADLCKAF 56  
DB 1 MDKFWMTAAWGLCLQSLAHPHOQIDLNTCYRAGVFHYEKNRGYSISRTAADLCKAF 60  
QY 57 NSTLPTMAOMEKALSIGETCRVGFIEGHVVIPIRHNSICAANTGVYL-TSNTSOYD 115  
DB 61 NSTLPTMDQMKALSKGETCRVGFIEGHVVIPIRHNAICANHGVYLLVTSNTSHYD 120  
QY 116 TYCFNAPAPREDCTSVTDLNPAFDGEPITITIVNRDGRYVQKGYRTNPEDIYPSNPTD 175  
DB 121 TYCFNAPAPREDCTSVTDLNPSFDGPVITITIVNRDGRYSKGEYRTHOEDIDASNIID 180  
QY 176 DVSSGSSERSSTSGGYIFTFSTVHPIDEDSPWITDSTRIPATNMSSHTLOPT 235  
DB 181 DVSSGSTIER-STPESTILHTY-----LPE-----QPT 209  
QY 236 ANPNTGLVEDLDRGPILSMTTQGSNSQSFSTSHGLEEDKHPTSTLTSNRNDVTGGR 295  
DB 210 G-----DQDSSFIRSTLATRDSDS----- 229  
QY 296 RDPNHSBSTLLBGYTSYHTKESRTFIPYTSAKTGSFGYAVTVGDSNVNRLSLG 355  
DB 230 -----SKDSR----- 234

QY 356 DQDTHPSGSGHT-THGSESDHSGSQEGGANTTSGPIPTPOIPEWLLILASLALALI 414  
DB 225 -----GSSRTVTHGSESLAGHSSANQDGVTTTSGPMRPPQIPEWLLILASLALALI 286  
QY 415 LAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNSESSTPDQ 474  
DB 287 LAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNSESSTPDQ 346  
QY 475 FMTADETRNLQSYDMKIGV 493  
DB 347 FMTADETRNLQSYDMKIGV 365

## RESULT 5

US-09-751-708A-73  
; Sequence 73, Application US/09751708A  
; Publication No. US20030157113A1  
; GENERAL INFORMATION:  
; APPLICANT: TERMAN, David S  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
; FILE REFERENCE: 751708  
; CURRENT APPLICATION NUMBER: US/09/751,708A  
; CURRENT FILING DATE: 2002-10-15  
; PRIOR APPLICATION NUMBER: US 60/173,371  
; PRIOR FILING DATE: 1999-12-28  
; NUMBER OF SEQ ID NOS: 166  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 73  
; LENGTH: 365  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-751-708A-73

Query Match 52.2%; Score 1361; DB 10; Length 365;  
Best Local Similarity 57.7%; Pred. No. 2.8e-100;  
Matches 288; Conservative 28; Mismatches 43; Indels 140; Gaps 9;

QY 1 MDKFWMTAAWGLCLVPLSLA-----QIDLNTTCRFAGVFHYEKNRGYSISRTAADLCKAF 56  
DB 1 MDKFWMTAAWGLCLQSLAHPHOQIDLNTCYRAGVFHYEKNRGYSISRTAADLCKAF 60  
QY 57 NSTLPTMAOMEKALSIGETCRVGFIEGHVVIPIRHNSICAANTGVYL-TSNTSOYD 115  
DB 61 NSTLPTMDQMKALSKGETCRVGFIEGHVVIPIRHNAICANHGVYLLVTSNTSHYD 120  
QY 116 TYCFNAPAPREDCTSVTDLNPAFDGEPITITIVNRDGRYVQKGYRTNPEDIYPSNPTD 175  
DB 121 TYCFNAPAPREDCTSVTDLNPSFDGPVITITIVNRDGRYSKGEYRTHOEDIDASNIID 180  
QY 176 DVSSGSSERSSTSGGYIFTFSTVHPIDEDSPWITDSTRIPATNMSSHTLOPT 235  
DB 181 DVSSGSTIER-STPESTILHTY-----LPE-----QPT 209  
QY 236 ANPNTGLVEDLDRGPILSMTTQGSNSQSFSTSHGLEEDKHPTSTLTSNRNDVTGGR 295  
DB 210 G-----DQDSSFIRSTLATRDSDS----- 229  
QY 296 RDPNHSBSTLLBGYTSYHTKESRTFIPYTSAKTGSFGYAVTVGDSNVNRLSLG 355  
DB 230 -----SKDSR----- 234  
QY 356 DQDTHPSGSGHT-THGSESDHSGSQEGGANTTSGPIPTPOIPEWLLILASLALALI 414  
DB 225 -----GSSRTVTHGSESLAGHSSANQDGVTTTSGPMRPPQIPEWLLILASLALALI 286  
QY 415 LAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNSESSTPDQ 474  
DB 287 LAVCIANVRRCQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVKNSESSTPDQ 346  
QY 475 FMTADETRNLQSYDMKIGV 493  
DB 347 FMTADETRNLQSYDMKIGV 365

RESULT 6  
US-10-291-634-3  
; Sequence 3, Application US/10291634  
; Publication No. US20030105058A1  
; GENERAL INFORMATION:  
; APPLICANT: Gentz, Reiner L  
; APPLICANT: Ni, Jian  
; APPLICANT: Dillon, Patrick J  
; TITLE OF INVENTION: CD-44 Like Protein  
; FILE REFERENCE: 1488, 0490003  
; CURRENT APPLICATION NUMBER: US/10/291,634  
; CURRENT FILING DATE: 2002-11-12  
; PRIOR APPLICATION NUMBER: 09/288,230  
; PRIOR FILING DATE: 1999-04-08  
; PRIOR APPLICATION NUMBER: 08/892,880  
; PRIOR FILING DATE: 1997-07-15  
; PRIOR APPLICATION NUMBER: 60/021,762  
; PRIOR FILING DATE: 1996-07-15  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 339  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-10-291-634-3  
Query Match 49.8%; Score 1299; DB 14; Length 339;  
Best Local Similarity 56.4%; Pred. No. 2.3e-95;  
Matches 279; Conservative 18; Mismatches 38; Indels 160; Gaps 8;  
QY 1 MDKFWHAANG-ICLVLSLA--QIDNITCRFAGVHHKNGKRSISRTAADLCKAN 57  
DB 1 MDKFWHAANG-ICLVLSLA--QIDNITCRFAGVHHKNGKRSISRTAADLCKAN 60  
QY 58 STPTMAQMEKALSIGETCRGYFIEGHVVIPIHPSICANNVTGYI-LTSNTSQYOT 116  
DB 61 TITPTMAQMEKALRKGETCRGYFIEGHVVIPIHPSICANNVTGYIILASNTSHYT 120  
QY 117 YCFNASAPBEDCTSVTDLPNADGPITITIVNRDGRYVQKGYRNPEDIYPSNPTD 176  
DB 121 YCFNASAPBEDCTSVTDLPNADGPITITIVNRDGRYVQKGYRNPEDIYPSNPTD 180  
QY 177 DVSSGSSERSSTGCIYFTFTSVHPIPEDCSPMITDSRIDRIPANMSSHTLQPTA 236  
DB 181 DVSSGSSERSSTGCIYFTFTSVHPIPEDCSPMITDSRIDRIPANMSSHTLQPTA 208  
QY 237 NPMTGLVEDLDRTGPLSMITQOQNSQSFSTSHGLEBKDHPTTSTLSSNRNDVTGGR 296  
DB 209 -----DUP-----TSQPT- 211  
QY 297 DPNHSGSTTLLEGYTHYPTKESRTFIPVTSKATGSPGVTAIVVGDNSNVNRLSGD 356  
DB 212 DDAFFIGSTL----- 221  
QY 357 QDTFHPGSGHTTHGSEDSHSGOEGANTTSGPIRTPOIEMIIILASLALALILA 416  
DB 222 -----ATGHSSGNQDSGVTTTSGPARRQIPEMIIILASLALALILA 264  
QY 417 VCIAYNSRRRCQKQKIVINSNGAVEDRPSGLNGEASQSEMVLVKNKESSETPDQM 476  
DB 265 VCIAYNSRRRCQKQKIVINSNGAVEDRPSGLNGEASQSEMVLVKNKESSETPDQM 324  
QY 477 TADETRVLQAVDMKI 491  
DB 325 TADETRVLQAVDMKI 339

RESULT 7  
US-10-156-932-18  
; Sequence 18, Application US/10156932  
; Publication No. US20030069181A1

GENERAL INFORMATION:  
; APPLICANT: Wong, Albert J.  
; TITLE OF INVENTION: Alternative Splice Forms of Proteins as  
; FILE REFERENCE: 8321-81  
; CURRENT APPLICATION NUMBER: US/10/156,932  
; CURRENT FILING DATE: 2002-05-28  
; PRIOR APPLICATION NUMBER: US 60/293,791  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 143  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-156-932-18  
Query Match 28.5%; Score 744; DB 14; Length 143;  
Best Local Similarity 100.0%; Pred. No. 1.3e-51;  
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 217 DRIPATNDSHSTTLQPTANPTGLVEDLDRTGPLSMITQOQNSQSFSTSHGLEBKD 276  
DB 1 DRIPATNDSHSTTLQPTANPTGLVEDLDRTGPLSMITQOQNSQSFSTSHGLEBKD 60  
QY 277 HPTTSTLSSNRNDVTGGRDPNHSSEGSTTLLEGYTHYPTKESRTFIPVTSKATGSPG 336  
DB 61 HPTTSTLSSNRNDVTGGRDPNHSSEGSTTLLEGYTHYPTKESRTFIPVTSKATGSPG 120  
QY 337 VTAVTVGDSNSNVNRLSGDQDT 359  
DB 121 VTAVTVGDSNSNVNRLSGDQDT 143  
RESULT 8  
US-10-156-932-20  
; Sequence 20, Application US/10156932  
; Publication No. US20030069181A1  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Albert J.  
; TITLE OF INVENTION: Alternative Splice Forms of Proteins as  
; FILE REFERENCE: 8321-81  
; CURRENT APPLICATION NUMBER: US/10/156,932  
; CURRENT FILING DATE: 2002-05-28  
; PRIOR APPLICATION NUMBER: US 60/293,791  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20  
; LENGTH: 185  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-156-932-20  
Query Match 28.2%; Score 734.5; DB 14; Length 185;  
Best Local Similarity 76.0%; Pred. No. 1.1e-50;  
Matches 149; Conservative 9; Mismatches 15; Indels 23; Gaps 2;  
QY 176 DVSSGSSERSSTGCIYFTFTSVHPIPEDS-----PWIDSDTRIPATN 223  
DB 1 DRIPATNDSHSTTLQPTANPTGLVEDLDRTGPLSMITQOQNSQSFSTSHGLEBKD 49  
QY 224 MDSSHSTTLQPTANPTGLVEDLDRTGPLSMITQOQNSQSFSTSHGLEBKD 283  
DB 50 MDSSHSTTLQPTANPTGLVEDLDRTGPLSMITQOQNSQSFSTSHGLEBKD 109  
QY 284 TSSNRNDVTGGRDPNHSSEGSTTLLEGYTHYPTKESRTFIPVTSKATGSPGVTAIVG 343  
DB 110 TSSNRNDVTGGRDPNHSSEGSTTLLEGYTHYPTKESRTFIPVTSKATGSPGVTAIVG 169  
QY 344 DSNENVRSLSGDQDT 359

Db 170 DSNNSNNRSLSGDQT 185

## RESULT 9

US-10-156-932-22  
; Sequence 22, Application US/10156932  
; Publication No. US2003006918A1  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Albert J.  
; TITLE OF INVENTION: Alternative Splice Forms of Proteins as  
; FILE REFERENCE: 8321-81  
; CURRENT APPLICATION NUMBER: US/10/156,932  
; PRIOR FILING DATE: 2002-05-28  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 187  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-156-932-22

Query Match 27.5%; Score 718; DB 14; Length 187;  
Best Local Similarity 82.5%; Pred. No. 2,2e-49;  
Matches 146; Conservative 3; Mismatches 10; Indels 18; Gaps 3;

QY 200 TVHPI-----PDDSPWINDSDRIP-----ATNMDSSTTLQPTANETGL 242  
DB 12 TSHWQGRTPSPEDSW-TDFVPLSHPMGRGHQGRMDSSHTILOPTANETGL 70  
QY 243 VEDLDRTGPLSMTQGSNSQSFSTSHGLEEDKDHPTSTLTSSNRDVTGGRDPNHS 302  
DB 71 VEDLDRTGPLSMTQGSNSQSFSTSHGLEEDKDHPTSTLTSSNRDVTGGRDPNHS 130  
QY 303 GSTTLLEGYTHSHYHTESRTFIPVTSKTSFGVTVVGVGDSNNRSLSGDQT 359  
DB 131 GSTTLLEGYTHSHYHTESRTFIPVTSKTSFGVTVVGVGDSNNRSLSGDQT 187

## RESULT 10

US-10-156-932-16  
; Sequence 16, Application US/10156932  
; Publication No. US2003006918A1  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Albert J.  
; TITLE OF INVENTION: Alternative Splice Forms of Proteins as  
; FILE REFERENCE: 8321-81  
; CURRENT APPLICATION NUMBER: US/10/156,932  
; PRIOR FILING DATE: 2002-05-28  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 109  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-156-932-16

Query Match 20.6%; Score 537; DB 14; Length 109;  
Best Local Similarity 99.0%; Pred. No. 2,9e-35;  
Matches 103; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 256 TQGSNSQSFSTSHGLEEDKDHPTSTLTSSNRDVTGGRDPNHSGSTTLLEGYTHY 315  
DB 6 TKGSNSQSFSTSHGLEEDKDHPTSTLTSSNRDVTGGRDPNHSGSTTLLEGYTHY 65  
QY 316 PHTKSRFTFIPVTSKTSFGVTVVGVGDSNNRSLSGDQT 359  
DB 66 PHTKSRFTFIPVTSKTSFGVTVVGVGDSNNRSLSGDQT 109

## RESULT 11

US-09-799-118-3  
; Sequence 3, Application US/09799118  
; Patent No. US20020090708A1  
; GENERAL INFORMATION:  
; APPLICANT: Lee, Tae Ho

Wislenski, Hans Georg  
Vilcek, Jan

TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding  
Thereof and Uses Thereof

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESS: Broadway and Neilmark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington

STATE: D.C.  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/799,118  
FILING DATE: 06-Mar-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/242,097  
FILING DATE: 13-MAY-1994

ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: LEE26/VILCEK=1B

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-628-5197  
TELEFAX: 212-737-3528

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 90 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-799-118-3

Query Match 18.3%; Score 478; DB 9; Length 90;  
Best Local Similarity 98.9%; Pred. No. 1,1e-30;  
Matches 89; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 32 GVEHYEKNGRYSISTEADICAKFNSLTPLMAQMERKLSIGFETCRGFTGHWVPIPI 91  
DB 1 GFVHEKNGRYSISTEADICAKFNSLTPLMAQMERKLSIGFETCRGFTGHWVPIPI 60  
QY 92 HENSICAAANTGVYILTSNTSQYDTYCFA 121  
DB 61 HENSICAAANTGVYILTSNTSQYDTYCFA 90

## RESULT 12

US-09-927-463-8  
; Sequence 8, Application US/09927463  
; Publication No. US20030032621A1  
; GENERAL INFORMATION:  
; APPLICANT: I.N.S.E.R.M.  
; TITLE OF INVENTION: Means for regulating hematopoietic differentiation  
; FILE REFERENCE: 1113  
; CURRENT APPLICATION NUMBER: US/09/927,463  
; PRIOR FILING DATE: 2001-12-17  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1

QY 32 GVEHYEKNGRYSISTEADICAKFNSLTPLMAQMERKLSIGFETCRGFTGHWVPIPI 91  
DB 1 GFVHEKNGRYSISTEADICAKFNSLTPLMAQMERKLSIGFETCRGFTGHWVPIPI 60  
QY 92 HENSICAAANTGVYILTSNTSQYDTYCFA 121  
DB 61 HENSICAAANTGVYILTSNTSQYDTYCFA 90

```
; SEQ ID NO 8
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-927-463-8

Query Match
Best Local Similarity 18.1%; Score 473; DB 10; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.8e-30;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LCLVPLSLAQIDINITCRFAGVFRVKNGRYSISRTEADLDCKAFSTPTMAQWEKALS 71
DB 1 LCLVPLSLAQIDINITCRFAGVFRVKNGRYSISRTEADLDCKAFSTPTMAQWEKALS 60

QY 72 IGFETCRYGFIIEGHVVIPIRIHPSICANN 101
DB 61 IGFETCRYGFIIEGHVVIPIRIHPSICANN 90

RESULT 13
US-10-156-932-14
; Sequence 14, Application US/10156932
; Publication No. US2003006916A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Albert J.
; TITLE OF INVENTION: Alternative Splice Forms of Proteins as
; FILE REFERENCE: 8321-81
; CURRENT APPLICATION NUMBER: US/10/156,932
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: US 60/293,791
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-156-932-14

Query Match
Best Local Similarity 14.8%; Score 386; DB 14; Length 79;
Best Local Similarity 94.8%; Pred. No. 2e-23;
Matches 73; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 283 LITSNRDVYGRGRRDPHSGSTTLLEGYTSHPHTKESRPTPTSAKTSFGYTAATV 342
DB 3 LITSNRDVYGRGRRDPHSGSTTLLEGYTSHPHTKESRPTPTSAKTSFGYTAATV 62

QY 343 GDSNSVNRSLSGDQDT 359
DB 63 GDSNSVNRSLSGDQDT 79

RESULT 14
US-10-052-641-2
; Sequence 2, Application US/10052641
; Publication No. US20030032073A1
; GENERAL INFORMATION:
; APPLICANT: HEIDER, et al.
; TITLE OF INVENTION: Method for Diagnosis and Therapy of Hodgkin Lymphomas
; FILE REFERENCE: 0652.1910000/RSP/KVY
; CURRENT APPLICATION NUMBER: US/10/052,641
; CURRENT FILING DATE: 2002-01-23
; PRIOR APPLICATION NUMBER: 09/331,254
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: PCT/EP97/07081
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: DE 196 53 607.3
; PRIOR FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 67
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-052-641-2

Query Match
Best Local Similarity 13.4%; Score 349; DB 14; Length 67;
Best Local Similarity 100.0%; Pred. No. 1.4e-20;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 RNDVYGRGRRDPHSGSTTLLEGYTSHPHTKESRPTPTSAKTSFGYTAATVGDNS 347
DB 1 RNDVYGRGRRDPHSGSTTLLEGYTSHPHTKESRPTPTSAKTSFGYTAATVGDNS 60

QY 348 NVNRSLS 354
DB 61 NVNRSLS 67

RESULT 15
US-09-909-320-201
; Sequence 201, Application US/09909320
; Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Batton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie F.
; APPLICANT: Pan, James
; APPLICANT: Paonli, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
```



pct-us01-51014-1.rapb

Page 8

Query Match	7.6%;	Score 199.5;	DB 9;	Length 322;
Best Local Similarity	22.2%;	Pred. No. 1.1e-07;		
Matches	83;	Conservative	48;	Mismatches 138;
			Indels	105;
			Gaps	12;

```

Oy 17 LSLAQDINTNCFACVFEVFNKGRYSISRLREADLCAFNSTLPTVAQNEKALSIGFET 76
Db 25 LRABELSTQVSCRIMITLIVSKANQOQMFPEAKACRLGLSLAGDQVETLAKXSFET 84
Oy 77 CRYGFT EGHVVI PRHENSICANNNGVYILTSNTS QYDTYCFAASAPBEDCISVTD 134
Db 85 CSYGWQDGGVVVISIRISPMKCKGRKGVGLIMKVYVSKQFAPACYNSSDWTNSCI---- 140
Oy 135 LPAAFDGPITITIVNDGTRVYQKEBYRNPEDIAPSPNDDDVVSSSSSESRSTSGYI 194
Db 141 -----PILITTKOPIFNTOJATQTHE-----FI 163
Oy 195 FYTFSTVHPIDEDSPMIDSTDRJPAITMDSHSTTLQTPANPNTGLAVEDLDRGTGLSM 254
Db 164 -----VSDST-----YSVASPYSTIAPFTTTPPAASISIRRRKCLIC 201
Oy 255 TIOO-SNSQSFSTSHGLEBDXDHPTTSTLTSSNNENDVYGRADPNHSEGSTLL----- 308
Db 202 VTEFVMEFSTMSIETEPVEENK-----AAFNBEAAGGVGP-----TALLVALL 246
Oy 309 -----EGTSHYPTKESRTPIPTSAKTSFGVTAIVYGDGNSNVNNSLSG 355
Db 247 FFGAALGLGCYVRYRVKAFPPFNKNQOKEMLEIYV-----VKEEKANDSNFN-ESEKTI 300
Oy 356 DODTFHPSGGSHTT 369
Db 301 DKNPBESKSPSKTT 314

```

Search completed: March 8, 2004, 06:22:12  
Job time : 398 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 8, 2004, 06:08:16 ; Search time 23 Seconds  
(without alignments)

1106.591 Million cell updates/sec

Title: PCT-US01-51014-1  
Perfect score: 2608  
Sequence: 1 MDKFWHAAWGLVPLSLA.....QFMADETRLNQVDMKIGV 493

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCITUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1827	70.1	361	1	US-07-946-497-6 Sequence 6, Appli
2	1827	70.1	361	1	US-08-483-322-6 Sequence 6, Appli
3	1827	70.1	361	2	US-08-478-882-6 Sequence 6, Appli
4	1698.5	65.1	362	6	5504194-2 Patent No. 5504194
5	1425	54.6	503	1	US-07-946-497-2 Sequence 2, Appli
6	1425	54.6	503	1	US-08-483-322-2 Sequence 2, Appli
7	1425	54.6	503	2	US-08-478-882-2 Sequence 2, Appli
8	1357	52.0	363	1	US-07-946-497-7 Sequence 7, Appli
9	1357	52.0	363	1	US-08-483-322-7 Sequence 7, Appli
10	1357	52.0	363	2	US-08-478-882-7 Sequence 7, Appli
11	1299	49.8	339	2	US-08-892-880-3 Sequence 3, Appli
12	711	27.3	354	1	US-07-946-497-4 Sequence 4, Appli
13	711	27.3	354	1	US-08-483-322-4 Sequence 4, Appli
14	711	27.3	354	2	US-08-478-882-4 Sequence 4, Appli
15	699	26.8	338	2	US-08-359-850-2 Sequence 2, Appli
16	478	18.3	90	2	US-08-242-097-3 Sequence 3, Appli
17	478	18.3	90	2	US-09-206-695-3 Sequence 3, Appli
18	478	18.3	90	4	US-09-799-118-3 Sequence 5, Appli
19	373	14.3	355	1	US-07-946-497-5 Sequence 5, Appli
20	373	14.3	355	1	US-08-483-322-5 Sequence 5, Appli
21	373	14.3	355	1	US-08-478-882-5 Sequence 5, Appli
22	362	13.9	334	2	US-08-359-850-4 Sequence 4, Appli
23	349	13.4	67	4	US-09-331-254-2 Sequence 2, Appli
24	226.5	8.7	41	2	US-08-143-311B-25 Sequence 25, Appli
25	209	8.0	42	2	US-08-143-311B-4 Sequence 4, Appli
26	209	8.0	42	4	US-08-753-851-4 Sequence 4, Appli
27	200	7.7	318	4	US-09-724-864-60 Sequence 60, Appli

28	199.5	7.6	332	4	US-09-232-160-21 Sequence 21, Appli
29	199.5	7.6	332	4	US-09-907-794A-201 Sequence 201, App
30	199.5	7.6	332	4	US-09-805-125A-201 Sequence 201, App
31	199.5	7.6	332	4	US-09-802-775A-201 Sequence 201, App
32	193.5	7.4	322	2	US-08-892-880-2 Sequence 9, Appli
33	187	7.2	78	3	US-08-564-225-9 Sequence 8, Appli
34	176	6.7	34	3	US-08-564-225-8 Sequence 8, Appli
35	174	6.7	36	2	US-08-143-311B-8 Sequence 3, Appli
36	170	6.5	32	2	US-08-143-311B-3 Sequence 3, Appli
37	170	6.5	32	4	US-08-753-851-3 Sequence 4, Appli
38	167.5	6.4	2137	4	US-09-134-001C-4463 Sequence 10, Appli
39	164	6.3	31	2	US-08-143-311B-10 Sequence 12, Appli
40	164	6.3	31	4	US-08-753-851-12 Sequence 2, Appli
41	162.5	6.2	277	1	US-08-024-868-2 Sequence 2, Appli
42	162.5	6.2	277	2	US-08-242-097-2 Sequence 2, Appli
43	162.5	6.2	277	3	US-09-206-695-2 Sequence 1, Appli
44	162.5	6.2	277	4	US-09-000-179-1 Sequence 2, Appli
45	162.5	6.2	277	4	US-09-799-118-2 Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-07-946-497-6  
Sequence 6, Application US/07946497  
Patent No. 5506119  
GENERAL INFORMATION:  
APPLICANT: HERRLICH, Peter  
APPLICANT: PONTA, Helmut  
APPLICANT: GUENTHER, Ursula  
APPLICANT: MATZKU, Siegfried  
APPLICANT: WENZL, Achim  
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA  
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,  
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/946,497  
FILING DATE: 19921109  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16915/145  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 361 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: hCD44  
US-07-946-497-6  
Query Match 70.1%; Score 1827; DB 1; Length 361;  
Best Local Similarity 72.6%; Pred. No. 3.6e-156;  
Matches 358; Conservative 0; Mismatches 3; Indels 132; Gaps 1;

```

QY 1 MDKFWHAAWGLCLVPLSLAQIDINTCRFAGVFEVKNGRYSISRTAADLCKAFNSTL 60
DB 1 MDKFWHAAWGLCLVPLSLAQIDINTCRFAGVFEVKNGRYSISRTAADLCKAFNSTL 60
QY 61 PTMAQMEKALSIGETCRGYFIEGHVVIPIRHPNSICAAANTGVYLLTNTSQDYTYCFN 120
DB 61 PTMAQMEKALSIGETCRGYFIEGHVVIPIRHPNSICAAANTGVYLLTNTSQDYTYCFN 120
QY 121 ASAPPEEDCTSVTDLFNAFDGPITITTVNRDGRTRYOKGEYRINPEDIYPSNPTDDVSS 180
DB 121 ASAPPEEDCTSVTDLFNAFDGPITITTVNRDGRTRYOKGEYRINPEDIYPSNPTDDVSS 180
QY 181 GSSSERSTSGGYFYFTSTVHPIDPDESPMITDSTDRIPATNMDSHSTTLQPTANPNT 240
DB 181 GSSSERSTSGGYFYFTSTVHPIDPDESPMITDSTDRIPATNMDSHSTTLQPTANPNT 240
QY 241 GLVEDLDRGTPLSMVTQOQNSQSFSTSHGLEEDKDHPTTSTLTSSNRNDVTGGRDPNH 300
DB 241 GLVEDLDRGTPLSMVTQOQNSQSFSTSHGLEEDKDHPTTSTLTSSNRNDVTGGRDPNH 300
QY 224 ----- 223
DB 224 ----- 223
QY 301 SEGSTTLLEGYTSHPHTKESRTFIPVTSAKTSGFVTAVTGDSNSNVNRSLSGDQDTF 360
DB 224 -----DQDTF 228
QY 361 HPSGGSHTTHGSESDGSHSGSQEGANTTSGPIRTPOIPFWLLILASLALAILAVCIA 420
DB 229 HPSGGSHTTHGSESDGSHSGSQEGANTTSGPIRTPOIPFWLLILASLALAILAVCIA 288
QY 421 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKQCEMVHLVNXESSSTPDPQFTADE 480
DB 289 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKQCEMVHLVNXESSSTPDPQFTADE 348
QY 481 TRNLQNVDMKIGV 493
DB 349 TRNLQNVDMKIGV 361

```

## RESULT 2

US-08-483-322-6

Sequence 6, Application US/08483322

Patent No. 5760178

GENERAL INFORMATION:

APPLICANT: HERRLICH, Peter

APPLICANT: PONTA, Helmut

APPLICANT: GUENTHER, Ursula

APPLICANT: MATZKU, Siegfried

APPLICANT: WENZL, Achim

TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Foley &amp; Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,322

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 07/946,497

FILING DATE: 09-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 16915/145

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 361 amino acids

TYPE: amino acid

TOPOLOGY: 1linear

IMMEDIATE SOURCE:

CLONE: hCD44

US-08-483-322-6

Query Match 70.1%; Score 1827; DB 1; Length 361;  
 Best Local Similarity 72.6%; Pred. No. 3; 6e-156;  
 Matches 358; Conservative 0; Mismatches 3; Indels 132; Gaps 1;

```

QY 1 MDKFWHAAWGLCLVPLSLAQIDINTCRFAGVFEVKNGRYSISRTAADLCKAFNSTL 60
DB 1 MDKFWHAAWGLCLVPLSLAQIDINTCRFAGVFEVKNGRYSISRTAADLCKAFNSTL 60
QY 61 PTMAQMEKALSIGETCRGYFIEGHVVIPIRHPNSICAAANTGVYLLTNTSQDYTYCFN 120
DB 61 PTMAQMEKALSIGETCRGYFIEGHVVIPIRHPNSICAAANTGVYLLTNTSQDYTYCFN 120
QY 121 ASAPPEEDCTSVTDLFNAFDGPITITTVNRDGRTRYOKGEYRINPEDIYPSNPTDDVSS 180
DB 121 ASAPPEEDCTSVTDLFNAFDGPITITTVNRDGRTRYOKGEYRINPEDIYPSNPTDDVSS 180
QY 181 GSSSERSTSGGYFYFTSTVHPIDPDESPMITDSTDRIPATNMDSHSTTLQPTANPNT 240
DB 181 GSSSERSTSGGYFYFTSTVHPIDPDESPMITDSTDRIPATNMDSHSTTLQPTANPNT 240
QY 241 GLVEDLDRGTPLSMVTQOQNSQSFSTSHGLEEDKDHPTTSTLTSSNRNDVTGGRDPNH 300
DB 241 GLVEDLDRGTPLSMVTQOQNSQSFSTSHGLEEDKDHPTTSTLTSSNRNDVTGGRDPNH 300
QY 224 ----- 223
DB 224 ----- 223
QY 301 SEGSTTLLEGYTSHPHTKESRTFIPVTSAKTSGFVTAVTGDSNSNVNRSLSGDQDTF 360
DB 224 -----DQDTF 228
QY 361 HPSGGSHTTHGSESDGSHSGSQEGANTTSGPIRTPOIPFWLLILASLALAILAVCIA 420
DB 229 HPSGGSHTTHGSESDGSHSGSQEGANTTSGPIRTPOIPFWLLILASLALAILAVCIA 288
QY 421 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKQCEMVHLVNXESSSTPDPQFTADE 480
DB 289 VNSRRRCQKKKLVINSNGAVEDRKPSGLNGEASKQCEMVHLVNXESSSTPDPQFTADE 348
QY 481 TRNLQNVDMKIGV 493
DB 349 TRNLQNVDMKIGV 361

```

## RESULT 3

US-08-478-882-6

Sequence 6, Application US/08478822

Patent No. 5885575

GENERAL INFORMATION:

APPLICANT: HERRLICH, Peter

APPLICANT: PONTA, Helmut

APPLICANT: GUENTHER, Ursula

APPLICANT: MATZKU, Siegfried

APPLICANT: WENZL, Achim

TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Foley &amp; Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

```

; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,882
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/946,497
; FILING DATE: 19921109
; ATTORNEY/AGENT INFORMATION:
; NAME: BERT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16915/145
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ. ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: hcd44
; US-08-478-882-6

```

```

Query Match      70.1%; Score 1827; DB 2; Length 361;
Best Local Similarity 72.6%; Pred. No. 3,6e-156;
Matches 358; Conservative 0; Mismatches 3; Indels 122; Gaps 1;

```

```

QY 1 MCKFWMAWGLCLVPLSLAQIDLNTTCFAGVFEKNGKRSISRTAADLCKAFNSTL 60
DB 1 MCKFWMAWGLCLVPLSLAQIDLNTTCFAGVFEKNGKRSISRTAADLCKAFNSTL 60
QY 61 PTMAQWEKALSIGFETCRVGFIEGHVVIPIRIHNSICAANTGVYILTSNTSOYDYTCFN 120
DB 61 PTMAQWEKALSIGFETCRVGFIEGHVVIPIRIHNSICAANTGVYILTSNTSOYDYTCFN 120
QY 121 ASAPPEDCTSVTDLNPAFDGPITITVNRDGRVYQKGRYTNPEDIYPSNPTDDVSS 180
DB 121 ASAPPEDCTSVTDLNPAFDGPITITVNRDGRVYQKGRYTNPEDIYPSNPTDDVSS 180
QY 181 GSSSERSTSGGIYFTFSTVHPIDEDSPWITDSTRIPATNMDSSHSTLQPTANPNT 240
DB 181 GSSSERSTSGGIYFTFSTVHPIDEDSPWITDSTRIPATNMDSSHSTLQPTANPNT 240
QY 241 GLVEDLDRGTPLSMTTQGSNSQSFSTSHGLEEDKHPTTSLTSSNNDVTGGRDPNH 300
DB 241 GLVEDLDRGTPLSMTTQGSNSQSFSTSHGLEEDKHPTTSLTSSNNDVTGGRDPNH 300
QY 224 ----- 223
DB 224 ----- 223
QY 301 SEGSTTLLEGYSHYPTKESRTFIPVTSKTSFGVAVTVGDSNVNRSLSGDQPTF 360
DB 224 -----DQDTF 228
QY 361 HPSGSGHTTHGSESDSHSGSQEGANTTSGPIRTPOIPEWLIILASLLALALIAVCI 420
DB 229 HPSGSGHTTHGSESDSHSGSQEGANTTSGPIRTPOIPEWLIILASLLALALIAVCI 288
QY 421 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNKSESSETPDQMTAD 480
DB 289 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNKSESSETPDQMTAD 348
QY 481 ETNQLQVDMKIGV 493
DB 349 ETNQLQVDMKIGV 361

```

RESULT 4  
5504194-2

```

; Patent No. 5504194
; APPLICANT: ST. JOHN, THOMAS P.; GALLATIN, W. MICHAEL; IDZERDA,
; REJEAN
; TITLE OF INVENTION: LYMPHOCYTE ADHESION RECEPTOR FOR HIGH
; ENOTHIELIN, CD44
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,624
; FILING DATE: 15-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 628,646
; FILING DATE: 12-DEC-1990
; APPLICATION NUMBER: 325,224
; FILING DATE: 17-MAR-1989
; SEQ. ID NO: 2:
; LENGTH: 362
; 5504194-2

```

```

Query Match      65.1%; Score 1698.5; DB 6; Length 362;
Best Local Similarity 68.2%; Pred. No. 1.3e-144;
Matches 337; Conservative 7; Mismatches 17; Indels 133; Gaps 2;

```

```

QY 1 MCKFWMAWGLCLVPLSLAQIDLNTTCFAGVFEKNGKRSISRTAADLCKAFNSTL 60
DB 1 MCKFWMAWGLCLVPLSLAQIDLNTTCFAGVFEKNGKRSISRTAADLCKAFNSTL 60
QY 61 PTMAQWEKALSIGFETCRVGFIEGHVVIPIRIHNSICAANTGVYILTSNTSOYDYTCFN 120
DB 61 PTMAQWEKALSIGFETCRVGFIEGHVVIPIRIHNSICAANTGVYILTSNTSOYDYTCFN 120
QY 121 ASAPPEDCTSVTDLNPAFDGPITITVNRDGRVYQKGRYTNPEDIYPSNPTDDVSS 180
DB 121 ASAPPEDCTSVTDLNPAFDGPITITVNRDGRVYQKGRYTNPEDIYPSNPTDDVSS 180
QY 181 GSSSERSTSGGIYFTFSTVHPIDEDSPWITDSTRIPATNMDSSHSTLQPTANPNT 229
DB 181 GSSSERSTSGGIYFTFSTVHPIDEDSPWITDSTRIPATNMDSSHSTLQPTANPNT 224
QY 240 GLVEDLDRGTPLSMTTQGSNSQSFSTSHGLEEDKHPTTSLTSSNNDVTGGRDPNH 299
DB 225 ----- 224
QY 300 HPSGSGHTTHGSESDSHSGSQEGANTTSGPIRTPOIPEWLIILASLLALALIAVCI 359
DB 225 -----DQGA 228
QY 360 HPSGSGHTTHGSESDSHSGSQEGANTTSGPIRTPOIPEWLIILASLLALALIAVCI 419
DB 229 HPSGSGHTTHGSESDSHSGSQEGANTTSGPIRTPOIPEWLIILASLLALALIAVCI 288
QY 420 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNKSESSETPDQMTAD 479
DB 289 VNSRRRCGQKKLVINSNGAVEDRKPSGLNGEASKSQEMVHLVNKSESSETPDQMTAD 348
QY 480 ETNQLQVDMKIGV 493
DB 349 ETNQLQVDMKIGV 362

```

```

RESULT 5
US-07-946-497-2
; Sequence 2, Application US/07946497
; Patent No. 5506119
; GENERAL INFORMATION:
; APPLICANT: HERPILICH, Peter
; APPLICANT: GUNTHER, Ursula
; APPLICANT: GUNTHER, Ursula
; APPLICANT: MATZKU, Siegfried
; APPLICANT: WENZL, Achim
; TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
; TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
; TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
; NUMBER OF SEQUENCES: 8

```

**Tue Mar 9 08:08:41 2004**

**pct-us01-51014-1.ra1**

Page 4

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/946,497  
FILING DATE: 19921109  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16915/145  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ. ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-946-497-2

Query Match	54.6%	Score 1425	DB 1	Length 503
Best Local Similarity	58.7%	Pred. No. 9.6e-120		
Matches 315	Conservative 36	Mismatches 108	Indels 78	Gaps 16

  

QY	1	MDKFMHAAWG-LCLVPLSLA--QIDILNITCRFAGVHVEKNGRYSISRTAEADLCKAFN	57
Db	1	MDKVMHTMGLCLLQSLAQQIDILNITCRAGVHVEKNGRYSISRTAEADLCEAN	60
QY	58	STLPVAAQEKALISIGFETCRGFTLEGHVIPIRIHNSICANNVGYI-LTNSISQYDT	116
Db	61	TLTPVTAQKELALRKGFETCRGFTLEGHVIPIRIHNAICANNVGYIILLASNTSHYDT	120
QY	117	YCNNAAPPEEDCTSTYDLPNAFDPIITITVNDGRRYQKGEYRNEDYTPSPNPD	176
Db	121	YCNNAAPPEEDCTSTYDLPNSFDPIITITVNDGRRYSKSGEYRNHODIDASITIDE	180

```

QY 177 DVSSGSSSRSSSGGYIYTT-FSYVHPIPED-----SPINDST-----D 217
Db 181 DVSSGSSIER-STPEEYIIMTLDPFSGPTGRDAFFIGSTLATIATPVASHTKQOE 239
QY 218 RIRATMDSHSTLLGPLANPMTGLVEDLR--TGPLSMTQOONSQSPFATSHGLEDK 275
Db 240 RLOMNPINHSPEVLQOTTR-----MTIDINSTSAIGEMWTGPORPFNNHVEODEET 294
QY 276 DHFTETLLSNANDVTG-----GRDPMHSGSTTLLEGYV--SHYPTHE 320
Db 295 PHATSTTMADPNSTTEALATQKKKPFENMGKNPPTPSEDS-HVTEGTTASHANNHPQ 353
QY 321 SRTFIPVTSATKGSFGVATVTVGDSNVNNSLSGDQTF---HPGSGHTTHGSESQ 376
Db 354 RMT-----TQOQBVSMT---DFPDPISHMGQCHQ---ESIG 386
QY 377 HSHGBOGGANTTSGEIRTPQIPENLIIILASLALALILACIAVNNRRRCQKKLVIN 436
Db 387 HSSGNDQDSGVTTSISGARRRPQIPENLIIILASLALALILACIAVNNRRRCQKKLVIN 446
QY 437 SGNGAYEDRRPSCSLNGLASQEMNLVYNKSSSETPDQPFMTADETRLQVNDKICV 493
Db 447 SGNGAYEDRRPSCSLNGLASQEMNLVYNKSETEPDDPFMTADETRLQVNDKICV 503

```

## RESULT 6

```

1 US-08-483-322-2
2 : Sequence 2, Application US/08483322
3 : Patent No. 5760178
4 :
5 : GENERAL INFORMATION:
6 :
7 : APPLICANT: HERRLICH, Peter
8 :
9 : APPLICANT: PONTA, Helmut
10 :
11 : APPLICANT: GUENTHER, Ursula
12 :
13 : APPLICANT: MATZKU, Siegfried
14 :
15 : APPLICANT: WENZL, Achim
16 :
17 : TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
18 :
19 : TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
20 :
21 : TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
22 :
23 : NUMBER OF SEQUENCES: 8
24 :
25 : CORRESPONDENCE ADDRESSES:
26 :
27 : ADDRESSEE: Foley & Lardner
28 :
29 : STREET: 3000 K Street, N.W., Suite 500
30 :
31 : CITY: Washington, D.C.
32 :
33 : COUNTRY: USA
34 :
35 : ZIP: 20007-5109
36 :
37 : COMPUTER READABLE FORM:
38 :
39 : MEDIUM TYPE: Floppy disk
40 :
41 : COMPUTER: IBM PC compatible
42 :
43 : OPERATING SYSTEM: PC-DOS/MS-DOS
44 :
45 : SOFTWARE: Patent in Release #1.0, Version #1.25
46 :
47 : CURRENT APPLICATION DATA:
48 :
49 : APPLICATION NUMBER: US/08/483,322
50 :
51 : FILING DATE: 07-JUN-1995
52 :
53 : CLASSIFICATION: 435
54 :
55 : PRIOR APPLICATION DATA:
56 :
57 : APPLICATION NUMBER: US 07/946,497
58 :
59 : FILING DATE: 09-NOV-1992
60 :
61 : ATTORNEY/AGENT INFORMATION:
62 :
63 : NAME: BENT, Stephen A.
64 :
65 : REGISTRATION NUMBER: 29,768
66 :
67 : REFERENCE/DOCKET NUMBER: 16915/145
68 :
69 : TELECOMMUNICATION INFORMATION:
70 :
71 : TELEPHONE: (202) 672-5300
72 :
73 : TELEFAX: (202) 672-5399
74 :
75 : TELEX: 904136
76 :
77 : INFORMATION FOR SEQ ID NO: 2:
78 :
79 : SEQUENCE CHARACTERISTICS:
80 :
81 : LENGTH: 503 amino acids
82 :
83 : TYPE: amino acid
84 :
85 : TOPOLOGY: linear
86 :
87 : MOLECULE TYPE: protein
88 :
89 :
90 :
91 :
92 :
93 :
94 :
95 :
96 :
97 :
98 :
99 :
100 :
101 :
102 :
103 :
104 :
105 :
106 :
107 :
108 :
109 :
110 :
111 :
112 :
113 :
114 :
115 :
116 :
117 :
118 :
119 :
120 :
121 :
122 :
123 :
124 :
125 :
126 :
127 :
128 :
129 :
130 :
131 :
132 :
133 :
134 :
135 :
136 :
137 :
138 :
139 :
140 :
141 :
142 :
143 :
144 :
145 :
146 :
147 :
148 :
149 :
150 :
151 :
152 :
153 :
154 :
155 :
156 :
157 :
158 :
159 :
160 :
161 :
162 :
163 :
164 :
165 :
166 :
167 :
168 :
169 :
170 :
171 :
172 :
173 :
174 :
175 :
176 :
177 :
178 :
179 :
180 :
181 :
182 :
183 :
184 :
185 :
186 :
187 :
188 :
189 :
190 :
191 :
192 :
193 :
194 :
195 :
196 :
197 :
198 :
199 :
200 :
201 :
202 :
203 :
204 :
205 :
206 :
207 :
208 :
209 :
210 :
211 :
212 :
213 :
214 :
215 :
216 :
217 :
218 :
219 :
220 :
221 :
222 :
223 :
224 :
225 :
226 :
227 :
228 :
229 :
230 :
231 :
232 :
233 :
234 :
235 :
236 :
237 :
238 :
239 :
240 :
241 :
242 :
243 :
244 :
245 :
246 :
247 :
248 :
249 :
250 :
251 :
252 :
253 :
254 :
255 :
256 :
257 :
258 :
259 :
260 :
261 :
262 :
263 :
264 :
265 :
266 :
267 :
268 :
269 :
270 :
271 :
272 :
273 :
274 :
275 :
276 :
277 :
278 :
279 :
280 :
281 :
282 :
283 :
284 :
285 :
286 :
287 :
288 :
289 :
290 :
291 :
292 :
293 :
294 :
295 :
296 :
297 :
298 :
299 :
300 :
301 :
302 :
303 :
304 :
305 :
306 :
307 :
308 :
309 :
310 :
311 :
312 :
313 :
314 :
315 :
316 :
317 :
318 :
319 :
320 :
321 :
322 :
323 :
324 :
325 :
326 :
327 :
328 :
329 :
330 :
331 :
332 :
333 :
334 :
335 :
336 :
337 :
338 :
339 :
340 :
341 :
342 :
343 :
344 :
345 :
346 :
347 :
348 :
349 :
350 :
351 :
352 :
353 :
354 :
355 :
356 :
357 :
358 :
359 :
360 :
361 :
362 :
363 :
364 :
365 :
366 :
367 :
368 :
369 :
370 :
371 :
372 :
373 :
374 :
375 :
376 :
377 :
378 :
379 :
380 :
381 :
382 :
383 :
384 :
385 :
386 :
387 :
388 :
389 :
390 :
391 :
392 :
393 :
394 :
395 :
396 :
397 :
398 :
399 :
400 :
401 :
402 :
403 :
404 :
405 :
406 :
407 :
408 :
409 :
410 :
411 :
412 :
413 :
414 :
415 :
416 :
417 :
418 :
419 :
420 :
421 :
422 :
423 :
424 :
425 :
426 :
427 :
428 :
429 :
430 :
431 :
432 :
433 :
434 :
435 :
436 :
437 :
438 :
439 :
440 :
441 :
442 :
443 :
444 :
445 :
446 :
447 :
448 :
449 :
450 :
451 :
452 :
453 :
454 :
455 :
456 :
457 :
458 :
459 :
460 :
461 :
462 :
463 :
464 :
465 :
466 :
467 :
468 :
469 :
470 :
471 :
472 :
473 :
474 :
475 :
476 :
477 :
478 :
479 :
480 :
481 :
482 :
483 :
484 :
485 :
486 :
487 :
488 :
489 :
490 :
491 :
492 :
493 :
494 :
495 :
496 :
497 :
498 :
499 :
500 :
501 :
502 :
503 :
504 :
505 :
506 :
507 :
508 :
509 :
510 :
511 :
512 :
513 :
514 :
515 :
516 :
517 :
518 :
519 :
520 :
521 :
522 :
523 :
524 :
525 :
526 :
527 :
528 :
529 :
530 :
531 :
532 :
533 :
534 :
535 :
536 :
537 :
538 :
539 :
540 :
541 :
542 :
543 :
544 :
545 :
546 :
547 :
548 :
549 :
550 :
551 :
552 :
553 :
554 :
555 :
556 :
557 :
558 :
559 :
560 :
561 :
562 :
563 :
564 :
565 :
566 :
567 :
568 :
569 :
570 :
571 :
572 :
573 :
574 :
575 :
576 :
577 :
578 :
579 :
580 :
581 :
582 :
583 :
584 :
585 :
586 :
587 :
588 :
589 :
590 :
591 :
592 :
593 :
594 :
595 :
596 :
597 :
598 :
599 :
600 :
601 :
602 :
603 :
604 :
605 :
606 :
607 :
608 :
609 :
610 :
611 :
612 :
613 :
614 :
615 :
616 :
617 :
618 :
619 :
620 :
621 :
622 :
623 :
624 :
625 :
626 :
627 :
628 :
629 :
630 :
631 :
632 :
633 :
634 :
635 :
636 :
637 :
638 :
639 :
640 :
641 :
642 :
643 :
644 :
645 :
646 :
647 :
648 :
649 :
650 :
651 :
652 :
653 :
654 :
655 :
656 :
657 :
658 :
659 :
660 :
661 :
662 :
663 :
664 :
665 :
666 :
667 :
668 :
669 :
670 :
671 :
672 :
673 :
674 :
675 :
676 :
677 :
678 :
679 :
680 :
681 :
682 :
683 :
684 :
685 :
686 :
687 :
688 :
689 :
690 :
691 :
692 :
693 :
694 :
695 :
696 :
697 :
698 :
699 :
700 :
701 :
702 :
703 :
704 :
705 :
706 :
707 :
708 :
709 :
710 :
711 :
712 :
713 :
714 :
715 :
716 :
717 :
718 :
719 :
720 :
721 :
722 :
723 :
724 :
725 :
726 :
727 :
728 :
729 :
730 :
731 :
732 :
733 :
734 :
735 :
736 :
737 :
738 :
739 :
740 :
741 :
742 :
743 :
744 :
745 :
746
```

Query March 1	54.6%:	Score 1425;	DB 1.1;	length 503;
Best Local Similarity	58.7%:	Pred. No. 9.6e-120;		
Matches 315;	Conservative 36;	Mismatches 108;	Indels 78;	Gaps 16

  

Qy	1	MDKFWMAAAG-ICTVPSLA--QIDNITCRFPAQVPHVEKNGRYSISRTBADLCAAFN	57
Db	1	MDKFWMTHTAGLCLICLQSLAQOQIDNITCRYAGVPHVEKNGYSISRTBADLCAAFN	60
Qy	58	SLPPTAAQEKSLISGFETCRYGFIQGHVVIPIRHPNSICAAANTGYLI-LTSNTSQYDT	116
Db	61	TLIPTAQVETLARKGFETCRYGFIQGHVVIPIRHPNICAANTGYLIILASNTSHYDT	120
Qy	117	YCNNAAPPEEDCTSYTDLPNADFDPITITIVNEDGTRYOKGGRYRNPEDIVSNYTD	176
Db	121	YCNNAAPPEEDCTSYTDLPNSFDPPIVITIVNHDGRYRSKKGIRHQSDIDASNLIIDE	180
Qy	177	DVSGSSSESSSTSGGVIIFYT-FSTVHPIDPED-----SPWITDST---D	217
Db	181	DVSGSGSTIEK-STPEGVYLIHTDLPFSQPTGRDAAFIQSTLATIATTPWVSATKONQE	239
Qy	218	RIRATMDSHSTTLOPTANPWTGLVGLDLR--DGPLSMTTQGSNSQSFSSHGGLDEK	275
Db	240	RIGWNEIHSNPVGLDTTTR---MTIDIKRNSISAHGEMNTQEPQRPFNHXYQDEET	294

Tue Mar 9 08:08:41 2004

pct-us01-51014-1.ra1

Page 5

Db 295 PHATSTTWADPNSTTEEAATQKEKPFENEMOGKNPPTSEDS-HYEGTTASAHNNHPSQ 353  
Qy 321 SRPFIVTSAKTGSFGVTAIVTVDGSDNSNVNRSLSGDQTF---HPGSGSHTHGSESDG 376  
Db 354 RMT-----TQSGEDVSWT-----DFDPDPISHMGQGHOT---ESKG 386  
Qy 377 HSHSQEGGANTTSGPIRTPOIPFWLIIILASLALALIIAVCIANVSRRCQKKLVIN 436  
Db 387 HSSGNQDSGVTTTSGPARRPOIPFWLIIILASLALALIIAVCIANVSRRCQKKLVIN 446  
Qy 437 SGNGAVEDRKPSGLNGEASKSQEMVHLVNXSESTPDQPMTADETRLQNVDMKIGV 493  
Db 447 SGNGVIEDRKPSGLNGEASKSQEMVHLVNXKEPTETPDQPMTADETRLQNVDMKIGV 503

RESULT 7  
US-08-478-882-2  
Sequence 2, Application US/08478882  
Patent No. 5885575  
GENERAL INFORMATION:  
APPLICANT: HERRLICH, Peter  
APPLICANT: PONTA, Helmut  
APPLICANT: GUENTHERT, Ursula  
APPLICANT: MATZKU, Siegfried  
APPLICANT: WENZL, Achim  
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA  
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,  
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,882  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/946,497  
FILING DATE: 19921109  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16915/145  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-478-882-2

Query Match 54.6%, Score 1425, DB 2, Length 503;  
Best Local Similarity 58.7%, Pred. No. 9.6e-120;  
Matches 315; Conservative 36; Mismatches 108; Indels 78; Gaps 16;

Qy 1 MDKFWTHAANG-LCLVPLSLA--QIDNITCRFAGVHVENGRYSISRTEADLCRAFN 57  
Db 1 MDKFWTHAANG-LCLVPLSLA--QIDNITCRFAGVHVENGRYSISRTEADLCRAFN 60  
Qy 58 STLTMAOMKALISGFETCRYGIEGHVIVPRHNSICAAANTGYIT-LTENTGYDT 116

Db 61 TTEPTAQMELAKKEETCRKYGIEGHVIVPRHNSICAAANTGYIT-LTENTGYDT 120  
Qy 117 YCFNAPSAPREEDCTSVTLDPNAPFDGPTITITIVNRDGRVYQKEGRINPEDIVPSNPTD 176  
Db 121 YCFNAPSAPREEDCTSVTLDPNAPFDGPTITITIVNRDGRVYQKEGRINPEDIVPSNPTD 180  
Qy 177 DVSSGSSSEASSTSGGIIFT-ESTVHPIDED-----SPWITST-----D 217  
Db 181 DVSSGSSSEASSTSGGIIFT-ESTVHPIDED-----SPWITST-----D 239  
Qy 218 RIPATNMDSSHSTLQPTANPTGLVVDLDR--TGPLSMTQOQNSQSFSSTHSGLEEDK 275  
Db 240 RTQNNPHNSNEVLLQTTT-----MTDIDRNSISANGEMWTGEPQPFNNHXYQDEET 294  
Qy 276 DHPTSTLTSSNRDVTG-----GRDDPNHSGSTTLLEGYT--SHYPTKE 320  
Db 295 PHATSTTWADPNSTTEEAATQKEKPFENEMOGKNPPTSEDS-HYEGTTASAHNNHPSQ 353  
Qy 321 SRPFIVTSAKTGSFGVTAIVTVDGSDNSNVNRSLSGDQTF---HPGSGSHTHGSESDG 376  
Db 354 RMT-----TQSGEDVSWT-----DFDPDPISHMGQGHOT---ESKG 386  
Qy 377 HSHSQEGGANTTSGPIRTPOIPFWLIIILASLALALIIAVCIANVSRRCQKKLVIN 436  
Db 387 HSSGNQDSGVTTTSGPARRPOIPFWLIIILASLALALIIAVCIANVSRRCQKKLVIN 446  
Qy 437 SGNGAVEDRKPSGLNGEASKSQEMVHLVNXSESTPDQPMTADETRLQNVDMKIGV 493  
Db 447 SGNGVIEDRKPSGLNGEASKSQEMVHLVNXKEPTETPDQPMTADETRLQNVDMKIGV 503

RESULT 8  
US-07-946-497-7  
Sequence 7, Application US/07946497  
Patent No. 5506119  
GENERAL INFORMATION:  
APPLICANT: HERRLICH, Peter  
APPLICANT: PONTA, Helmut  
APPLICANT: GUENTHERT, Ursula  
APPLICANT: MATZKU, Siegfried  
APPLICANT: WENZL, Achim  
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA  
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,  
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/946,497  
FILING DATE: 19921109  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16915/145  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 363 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: MCD44  
US-07-946-497-7

Query Match 52.0%; Score 1357; DB 1; Length 363;  
Best Local Similarity 57.9%; Pred. No. 7.6e-114; Indels 138; Gaps 9;  
Matches 288; Conservative 28; Mismatches 43; Indels 138; Gaps 9;

1 MDKFWMAWAGLCLVPLSLA--QIDNITCRFAGVHVEKNRGYSISRTAADLCKAFNS 58  
1 MDKFWMAWAGLCLVPLSLA--QIDNITCRFAGVHVEKNRGYSISRTAADLCKAFNS 60  
59 TLPTMAQMEKALISGFTCRGFTGEGHVIPIRIHNSICAAANTGYIIL-TSNTSOYDTY 117  
61 TLPTMDQKALSKGFETCRGFTGEGHVIPIRIHNSICAAANTGYIILVTSNTSHDTY 120  
118 CFNASAPPEEDCTSVTDLFNAFDGPITITIVNRDGTTRYVQGEYKTNPEDIYPSNPTDD 177  
121 CFNASAPPEEDCTSVTDLFNSFDGPVITITIVNRDGTTRYVQGEYKTHQEDIDASNITDD 180  
178 VSSGSSSRSTSGYIYFTSTVHPIDEDSPWITDSTRIPATNMDSHSTLQPTAN 237  
181 VSSGSTIEK-STPEGYILHTY-----LPTB-----OPTG- 208  
238 PNTGLVEDLDRTGSLMTQOQNSQSFSTSHGLEEDKHPTTSTLSSNRNDVTGARD 297  
209 -----DQDPSFIRSTLATRDRS----- 227  
298 PNHSEGSTTLEGTSHYPTKESRTPIPTSAKTSFGVTAATVGDNSNVNRLSGDQ 357  
228 -----SKDSR----- 232  
358 DTFHPSGGSHT-TGSESDGSHSGOEGANTTSGPIRTPOIPEWIIILASLALALILA 416  
233 -----GSRRTVTHGSELAGHSANQDSGVTTTSGPMRRPOIPEWIIILASLALALILA 286  
417 VCIANVSRRCGQKKLVYNSGNAVEDRKPSGLNGEASKSQEMVHLVKNKESSETPDOFM 476  
287 VCIANVSRRCGQKKLVYNSGNAVEDRKPSGLNGEASKSQEMVHLVKNKESSETPDOCM 346  
477 TADETNLQNVDMKIGV 493  
347 TADETNLQNVDMKIGV 363

RESULT 9

US-08-483-322-7  
Sequence 7, Application US/08483322  
Patent No. 5760178  
GENERAL INFORMATION:  
APPLICANT: HERRLICH, Peter  
APPLICANT: PONTA, Helmut  
APPLICANT: GUENTHER, Ursula  
APPLICANT: MATZKU, Siegfried  
APPLICANT: WENZL, Achim  
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA  
TITLE OF INVENTION: SEQUENCES CODING THESE ANTIBODIES AGAINST THESE PROTEINS,  
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,322  
FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/946,497  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16915/145  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 363 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: MCD44  
US-08-483-322-7

Query Match 52.0%; Score 1357; DB 1; Length 363;  
Best Local Similarity 57.9%; Pred. No. 7.6e-114; Indels 138; Gaps 9;  
Matches 288; Conservative 28; Mismatches 43; Indels 138; Gaps 9;

1 MDKFWMAWAGLCLVPLSLA--QIDNITCRFAGVHVEKNRGYSISRTAADLCKAFNS 58  
1 MDKFWMAWAGLCLVPLSLA--QIDNITCRFAGVHVEKNRGYSISRTAADLCKAFNS 60  
59 TLPTMAQMEKALISGFTCRGFTGEGHVIPIRIHNSICAAANTGYIIL-TSNTSOYDTY 117  
61 TLPTMDQKALSKGFETCRGFTGEGHVIPIRIHNSICAAANTGYIILVTSNTSHDTY 120  
118 CFNASAPPEEDCTSVTDLFNAFDGPITITIVNRDGTTRYVQGEYKTNPEDIYPSNPTDD 177  
121 CFNASAPPEEDCTSVTDLFNSFDGPVITITIVNRDGTTRYVQGEYKTHQEDIDASNITDD 180  
178 VSSGSSSRSTSGYIYFTSTVHPIDEDSPWITDSTRIPATNMDSHSTLQPTAN 237  
181 VSSGSTIEK-STPEGYILHTY-----LPTB-----OPTG- 208  
238 PNTGLVEDLDRTGSLMTQOQNSQSFSTSHGLEEDKHPTTSTLSSNRNDVTGARD 297  
209 -----DQDPSFIRSTLATRDRS----- 227  
298 PNHSEGSTTLEGTSHYPTKESRTPIPTSAKTSFGVTAATVGDNSNVNRLSGDQ 357  
228 -----SKDSR----- 232  
358 DTFHPSGGSHT-TGSESDGSHSGOEGANTTSGPIRTPOIPEWIIILASLALALILA 416  
233 -----GSRRTVTHGSELAGHSANQDSGVTTTSGPMRRPOIPEWIIILASLALALILA 286  
417 VCIANVSRRCGQKKLVYNSGNAVEDRKPSGLNGEASKSQEMVHLVKNKESSETPDOFM 476  
287 VCIANVSRRCGQKKLVYNSGNAVEDRKPSGLNGEASKSQEMVHLVKNKESSETPDOCM 346  
477 TADETNLQNVDMKIGV 493  
347 TADETNLQNVDMKIGV 363

RESULT 10

US-08-478-882-7  
Sequence 7, Application US/08478882  
Patent No. 5885575  
GENERAL INFORMATION:  
APPLICANT: HERRLICH, Peter  
APPLICANT: PONTA, Helmut  
APPLICANT: GUENTHER, Ursula  
APPLICANT: MATZKU, Siegfried  
APPLICANT: WENZL, Achim  
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

Query Match	52.0%	Score 1357	DB 2	Length 369
Best Local Similarity	57.9%	Pred. No. 7.6e-114		
Matches 286	Conservative 28	Mismatches 43	Indels 138	Gaps 9
Qy	1	MDKFWMAHAAWGLCLVPLSLA--QIDLNTTCFACVCFHYEKXGRGRYSISRTLEADLCKAFNS	58	
Db	1	MDKFWHHTTAMGCLLTQSLSLAHQOIDLNTTCYACVCFYKXGRGRYSISRTLEADLCKAFNS	60	
Qy	59	TLPTMAQMKALSIGFTGCRGFIIRGHVIRIRHPSICAAANTGVYL--TSTTSQYDHY	117	
Db	61	TLPTMQMKALSKGFTGCRGFIIRGHVIRIRHPSICAAANTGVYILTSTSTSHDY	120	
Qy	118	CFNAPASRPEDCTSVYDLFNAFDGPIITITIVNRDQIRYVQKGEKRTNPEDYPSNPTDD	177	
Db	121	CFNAPASRPEDCTSVYDLFNSFDGPVITITIVNRDQIRYKXGEKRTHQEDIDASNITDD	180	
Qy	178	VSSGSSSESSNSGVIITFTFTSTVAPRIDEDSPMITDSTDIRAPATNDDSHSTTLOPTAN	237	
Db	181	VSSGSIYTER-STPEGILHTY-----LPT-----OPTG--	208	
Qy	238	PNTGLVEDLDRTGPLSMITTOGSNSQSFSTSHGLEDKHPHTTSLTNSNRNDVYTGARD	297	
Db	209	-----DQDQSPFIRSTLATRDSD-----	227	
Qy	298	PNHSBGSTTLLEGYTHRYHTKESRTFLPVTSAKTGSGCYATVAVGDSNSNVKRLSDGO	357	
Db	228	-----SKDSR-----	232	
Qy	358	DFEHPGGSHT--THGSESDGHGSGQOEGAGNTSGPIRTPIPMILITLALALALILA	416	
Db	233	-----GSSTVTHHSELHGSSANQDSGVTTTSGPMKRPQIPMWLITLALALALILA	286	
Qy	417	VCLAVNSRRRCQKKCLVTNSGNGAVEDRKPSGLNGEASKSQEWVHLVNTSESETDQM	476	
Db	287	VCLAVNSRRRCQKKCLVTNGANGVATEDKPSSELNGEASKSQEWVHLVNTSEPSITDQC	346	

237 NPNTGLVEDLDRIC QY

QY	1	MDKFWHAWNC-LCLVPISLA--QIDANTIRPAGVFVEVNGRISIRTAADICKFN	57
Dh	1	MDKWMHAWMGLLLOLSIAQOQIDNITIRYAGVFVEVNGRISIRTEAADICEAFN	60
QY	58	STLEPMAQWELALSIGPFCRYGFIIEGHVVIPIRHPSNICAANTGYVI-LTSNTSQYDT	116
Dh	61	TTLPTLMAQWELALKKGFCTCRYGFIIEGHVVIPIRHPSNICAANTGVILLASNTSHDT	120
QY	117	YCFNAPSAPBEDCSTVTDLPNAPGPIITITIVNBDGTYYVKGEXRINPEDIYPSNPYTD	176
Dh	121	YCFNAPSAPLEEDCSTVTDLPNPSFGOPVITITIVNRDGTYSKKGEYRTQEDIDASNIID	180
QY	177	DVSGGSSSEBSSTGGYIFTFEFSVWHPFPDSDSPWITDSTDIRIPATNMDSHSTLQETA	236
Dh	181	DVSGGSTIEK-STREGYILHNT-----DLP-----TSQPT-	208
QY	237	NPNTGCVEDDLRTGPILSMTTOQSNQSPTSISHEGLEEDKHPHTTSLTSSNRNDVTVGRR	296



Db 209 -----GDR 211  
QY 297 DPNHSGSTLLBEGYSHYPTKESRTFFIPVTSKATGSGVAVTVGDSNSNVNRLSGD 356  
Db 212 DDAFFIGSTL----- 221  
QY 357 ODFHPGSGSHHTHSGSDGSHSGOGGANTTSGPIRTPQIPENTIIILASLLATALLA 416  
Db 222 -----ATGHSNGNDSSGVTTSNGPARRQIPENIIILASLLATALLA 264  
QY 417 VCIANSRRRCGQKKKLVINSNGAVZDRKPSGLNGBASCSQMVHLVNKESSETPDQM 476  
Db 265 VCIANSRRRCGQKKKLVINSNGVTRKPSGLNGBASCSQMVHLVNKESSETPDQM 324  
QY 477 TADETRLQNVDMKI 491  
Db 325 TADETRLQNVDMKI 339

RESULT 12  
US-07-946-497-4  
; Sequence 4, Application US/07946497  
; Patent No. 5506119  
GENERAL INFORMATION:  
APPLICANT: HERRLICH, Peter  
APPLICANT: PONTA, Helmut  
APPLICANT: GUENTHERT, Ursula  
APPLICANT: MATZKU, Siegfried  
APPLICANT: WENZL, Achim  
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA  
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,  
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/946,497  
FILING DATE: 19921109  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16915/145  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 354 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-946-497-4

Query Match 27.3%; Score 711; DB 1; Length 354;  
Best Local Similarity 70.0%; Pred. No. 1,1e-55;  
Matches 149; Conservative 10; Mismatches 18; Indels 36; Gaps 5;  
QY 161 YRTNEDYIPSNPTDDVSSGSSRSSTSGGYIFRTSTVHPI-----PDDESPWITD 214  
Db 161 YRTNEDYIPSNPTDDVSSGSSRSSTSGGYIFRTSTVHPI-----PDDESPWITD 201  
QY 215 STDRIIP-----ATNMDSHSTLQPTANPNTGLVEDLDRGTGLSMWTOQSNOS 263

Db 202 FPNPISHPMGRGHQAGRBMWDSHSTLLQPTANPNTGLVEDLDRGTGLSMWTOQSNOS 261  
QY 264 FSTHGELEEDKHPTTSTLLTSNRNDVYTGRRDPNHSBEGSTLLBEGYSHYPTKESRT 323  
Db 262 FSTHGELEEDKHPTTSTLLTSNRNDVYTGRRDPNHSBEGSTLLBEGYSHYPTKESRT 321  
QY 324 FIPVTSKATGSGVAVTVGDSNSNVNRLSGD 356  
Db 322 FIPVTSKATGSGVAVTVGDSNSNVNRLSGD 354

RESULT 13  
US-08-483-322-4  
; Sequence 4, Application US/08483322  
; Patent No. 5760178  
GENERAL INFORMATION:  
APPLICANT: HERRLICH, Peter  
APPLICANT: PONTA, Helmut  
APPLICANT: GUENTHERT, Ursula  
APPLICANT: MATZKU, Siegfried  
APPLICANT: WENZL, Achim  
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA  
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,  
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,322  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/946,497  
FILING DATE: 09-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16915/145  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 354 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-483-322-4

Query Match 27.3%; Score 711; DB 1; Length 354;  
Best Local Similarity 70.0%; Pred. No. 1,1e-55;  
Matches 149; Conservative 10; Mismatches 18; Indels 36; Gaps 5;  
QY 161 YRTNEDYIPSNPTDDVSSGSSRSSTSGGYIFRTSTVHPI-----PDDESPWITD 214  
Db 161 YRTNEDYIPSNPTDDVSSGSSRSSTSGGYIFRTSTVHPI-----PDDESPWITD 201  
QY 215 STDRIIP-----ATNMDSHSTLQPTANPNTGLVEDLDRGTGLSMWTOQSNOS 263  
Db 202 FPNPISHPMGRGHQAGRBMWDSHSTLLQPTANPNTGLVEDLDRGTGLSMWTOQSNOS 261  
QY 264 FSTHGELEEDKHPTTSTLLTSNRNDVYTGRRDPNHSBEGSTLLBEGYSHYPTKESRT 323

Db 262 FSTSHGLEDKDHPTTSTLTSSNRNDVTGGRDPNHSSEGSTLLLEGYTSHPHTKESRT 321  
QY 324 FIPVTSKATGSFGVTAIVTGDNSNVNRSLSGD 356  
Db 322 FIPVTSKATGSFGVTAIVTGDNSNVNRSLSGD 354

RESULT 14  
US-08-478-882-4  
Sequence 4, Application US/08478882  
Patent No. 5885575  
GENERAL INFORMATION:  
APPLICANT: HERLICH, Peter  
APPLICANT: PONTA, Helmut  
APPLICANT: GUENTHER, Ursula  
APPLICANT: MATZKU, Siegfried  
APPLICANT: WENZL, Achim  
TITLE OF INVENTION: VARIANT CD4 SURFACE PROTEINS, DNA  
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,  
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,882  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/946,497  
FILING DATE: 19921109  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16915/145  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 354 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-478-882-4

Query Match 27.3%; Score 711; DB 2; Length 354;  
Best Local Similarity 70.0%; Pred. No. 1.1e-55;  
Matches 149; Conservative 10; Mismatches 18; Indels 36; Gaps 5;

QY 161 YRTNPEDLYPSNPDDDDVSSGSSERSSTSGGYFFYFSTVHPI-----PDEDSPWITD 214  
Db 161 YRQTPRE-----DSHSTGTAAASAHTS-----HPMQGRTPSPEDSSW-TD 201

QY 215 STDRIP-----ATNNDSSHSTTLQPTANPNTGLVEDLDRGTGLMTTQOOSNSQS 263  
Db 202 FFPNISHPMGRGHQGRMDSDSHSTTLQPTANPNTGLVEDLDRGTGLMTTQOOSNSQS 261

QY 264 FSTSHGLEDKDHPTTSTLTSSNRNDVTGGRDPNHSSEGSTLLLEGYTSHPHTKESRT 323  
Db 262 FSTSHGLEDKDHPTTSTLTSSNRNDVTGGRDPNHSSEGSTLLLEGYTSHPHTKESRT 321

QY 324 FIPVTSKATGSFGVTAIVTGDNSNVNRSLSGD 356

Db 322 FIPVTSKATGSFGVTAIVTGDNSNVNRSLSGD 354

RESULT 15  
US-08-359-850-2  
Sequence 2, Application US/08359850  
Patent No. 5951982  
GENERAL INFORMATION:  
APPLICANT: Ziller, Margot  
APPLICANT: Herlich, Peter  
APPLICANT: Ponta, Helmut  
TITLE OF INVENTION: Use of Antibody-Containing Preparations  
TITLE OF INVENTION: for Immunosuppression  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein and Fox  
STREET: 1225 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/359,850  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/963,323  
FILING DATE:  
APPLICATION NUMBER: DE P 41 34 982.2  
FILING DATE: 23-OCT-1991  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 833-7553  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 338 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-359-850-2

Query Match 26.8%; Score 699; DB 2; Length 338;  
Best Local Similarity 69.7%; Pred. No. 1.2e-54;  
Matches 147; Conservative 10; Mismatches 18; Indels 36; Gaps 5;

QY 161 YRTNPEDLYPSNPDDDDVSSGSSERSSTSGGYFFYFSTVHPI-----PDEDSPWITD 214  
Db 147 YRQTPRE-----DSHSTGTAAASAHTS-----HPMQGRTPSPEDSSW-TD 187

QY 215 STDRIP-----ATNNDSSHSTTLQPTANPNTGLVEDLDRGTGLMTTQOOSNSQS 263  
Db 188 FFPNISHPMGRGHQGRMDSDSHSTTLQPTANPNTGLVEDLDRGTGLMTTQOOSNSQS 247

QY 264 FSTSHGLEDKDHPTTSTLTSSNRNDVTGGRDPNHSSEGSTLLLEGYTSHPHTKESRT 323  
Db 248 FSTSHGLEDKDHPTTSTLTSSNRNDVTGGRDPNHSSEGSTLLLEGYTSHPHTKESRT 307

QY 324 FIPVTSKATGSFGVTAIVTGDNSNVNRSLS 354  
Db 308 FIPVTSKATGSFGVTAIVTGDNSNVNRSLS 338

Search completed: March 8, 2004, 06:10:38  
Job time : 26 secs